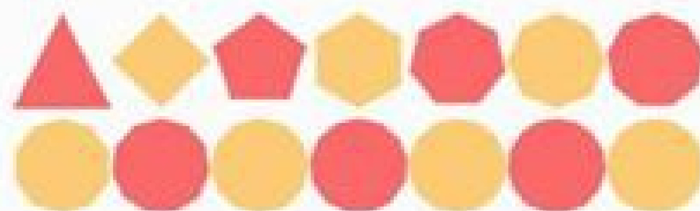




数字化城市

实时数据如何改变城市化进程



[DE]CODING THE CITY

How Real-Time Data Changes Urbanism

迪特马尔·奥夫胡贝尔
卡洛·拉蒂 主编

曹海元 常莹 陈然
贺霄羽 黄艺 姚琪 译



东西文库

目　录

编辑手记：迈向智能城市

青苹果数据中心
Green Apple Data Center

数字化城市

实时数据如何改变城市化进程

［美］迪特马尔·奥夫胡贝尔（**Dietmar Offenhuber**）、［美］卡洛·拉蒂（**Carlo Ratti**） 主编；曹海元、常莹、陈然、党霄羽、黄艺、姚琪 译

出版社：东西文库

湖南省青苹果数据中心有限公司

注册时间：1992年8月13日

注册地址：中国湖南省长沙市开福区青竹湖大道399号

互联网出版许可证：新出网证（湘）字013号

电子邮箱：GA@egreenapple.com

网　　址：www.egreenapple.com

青苹果数据中心为作者和相关机构提供数字出版服务。

本书电子版如有错讹，敬请指正，我们会及时更新版本。

关键词

数字化城市　智能城市　大数据　科技　城市化　东西文库

一句话宣传语

麻省理工大学可感城市实验室前沿成果

内容简介

《数字化城市》这本书是麻省理工（MIT）工大学可感城市实验室（Senseable City Lab）的一本论文集，也是东西文库的第一本公益电子书。这本书是“数字化城市”（或者用时髦一点的词“智能城市”）领域的一本重要的论文集，从城市数据的采集、表现、意义及应用等方面探讨数据对城市生活和城市发展的影响。

作者简介

卡洛·拉蒂，一位训练有素的建筑师、工程师，麻省理工大学教师，负责指导可感城市实验室的工作。他也是卡洛·拉蒂协会国际设计办公室的创始成员。他目前是世界经济论坛全球议程委员会（World Economic Forum Global Agenda Council）城市管理部门的成员，担任2015年世博会未来食品馆的馆长。拉蒂被快公司（Fast Company）名列为“美国50位最有影响力的设计师”，《连线》杂志名列为“50位改变世界的人物”。

迪特马尔·奥夫胡贝尔是美国东北大学艺术与设计和公共政策系的助理教授。他的研究关注在城市管理与市民话语中可视化与表象的作用，他负责许多关于基础设施的清晰度与垃圾处理系统的项目。奥夫胡贝尔是麻省理工大学城市研究与规划系的博士，维也纳科技大学与麻省理工大学媒体实验室的硕士。他获了好几个奖项，包括NSF可视化挑战、《美国规划协会》颁发的年度最佳论文奖。

译者简介

常莹，中央财经大学（Central University of Finance and Economics）宏观经济统计本科毕业，北京大学（Peking University）光华商务统计硕士，《为女权辩护》负责人、译者。《数字化城市》译者。

曹海元，复旦大学（Fudan University）物理学本科毕业，直博生在读，《波希米亚在伦敦》译者。《数字化城市》译者。

陈然，上海财经大学（Shanghai University of Finance and Economics）金融硕士，银行从业两年，取得过CATTI三级笔译证书。《数字化城市》译者。

党霄羽，华东师范大学（East China Normal University）中文系文艺学硕士，《夜贼莱佛士》负责人，《大学的兴起及其早期架构》负责人、译者，万古杂志中文项目负责人之一。《数字化城市》负责人、译者。

黄艺，加利福尼亚大学洛杉矶分校（University of California, Los Angeles, UCLA）机械工程和数学本科学位，麻省理工大学（Massachusetts Institute of Technology, MIT）机械工程硕士，机械工程博士。《数字化城市》译者。

姚琪，IT工程师，《FPGA快速系统原型设计权威指南》译者，现正参与翻译专业书籍《Embedded System Security: Practical Methods for Safe and Secure Software and Systems Development》。《数字化城市》译者。

微博段子

麻省理工大学可感城市实验室的研究成果，利用大数据对城市生活的各个方面进行了分析。

新浪微博ID

@cycufepku（常莹）　@跑在路上寻觅（姚琪）　@raincoatrun（曹海元）　@独眼一点五（党霄羽）@-Rane-（陈然）@Rita90star（黄艺）　@幽灵flying（丁诗颖）　@东西文库

中文版序（一）

作者　迪特马尔·奥夫胡贝尔　卡洛·拉蒂

译者　党霄羽

很高兴能为中文翻译版写这篇序。对于城市研究这门学科来说，21世纪是一个让人兴奋的时代——对全世界而言都是如此，而在东南亚让人尤为激动。这一地区极为活跃，目前的人口比全世界其他地方要多。城市化势头迅猛，特别是中国，在本世纪建造的城市肌理比此前所有人类建造的都多。

在上个世纪80年代的这场建造比赛的一开始，中国的主要精力还是放在追赶建筑与规划的先进实践方面。如今在城市化领域，这个国家正在把世界上的其他国家甩在身后，我们需要创新。毫无疑问，在未来，大数据将在城市化的创新中扮演至关重要的角色。我们怀着兴奋的心情，设想中国的创新者们、城市研究人员、建筑师、开发人员、数据科学家、设计师们会被我们这本书中勾勒的不同观点所启发，在未来对这些观点加以发展，在中国的大环境中取得丰硕的成果。

中文版序（二）

作者　宋朝鸣

2008年的初春，我从纽约搬到波士顿，加入东北大学（Northeastern University）的复杂网络研究中心（Center for Complex Network Research，CCNR）。这是一个志于研究不同领域网络结构以及其动力学的跨学科机构，其成员大多是来自于数学物理生物社会学计算科学等等不同背景的科学家，中心的负责人Albert-László Barabási是网络科学的先驱和奠基人之一。CCNR独占了大楼的顶层，向窗外远眺有相当不错的景致，其间从会议室到报告厅一应俱全，宛若一个迷你的研究所。那个时候整个研究小组刚从圣母大学（University of Notre Dame）迁来不久，大家都对这个的新成立不久的研究中心充满了期待。

当时中心一个重要的研究方向，是关于定量化认识人类流动规律，这是一个相当新颖而有趣的课题。仅仅两年前，Dirk Brockmann和他当时在美国西北大学的研究小组在《自然》上发表了一篇关于人类旅行标度律的文章［1］。其中分析的数据源自于wheresgeorge. com，这是一个追踪美国纸币在地理上流动方向的网站（其域名翻译成中文叫“乔治在哪里”，而一美元纸钞正面印的正是首任总统乔治华盛顿的头像）。通过众多网友自发上传自己所拥有的纸钞编号，Brockmann等人的工作重构了46万张纸币的移动轨迹，进而在连续时间随机游走（continuous time random walk）的数学框架下架构了一个模型，能在相当程度上刻画纸币流动的规律，这是第一次大规模对人类流动规律进行定量研究。随机游走模型一直在物理领域被广泛地使用，经常用来描述微观粒子在随机热扰动下的不规则运动轨迹。然而，相较于粒子运动相对稳定的迁移步长，Brockmann模型所描述的纸币流动步长存在着相当大的不确定性：在很多时候纸币只在很小的区域内流动，但是偶尔也会进行大尺度上的跳跃，具有所谓列维飞行（Lévy flight）的特征。不过，纸币并不

能长时间的停留在特定人手中，所以该模型只是对人类流动规律的一个粗糙近似。

而我们对这个问题感兴趣的原因在于，我们找到了更好的追踪器——手机。大约也在一两年前，研究中心开始跟欧洲的某电信公司进行研究合作，由他们提供将近一千万匿名手机用户的通讯数据。内容大致上包括了每个用户的给其他用户的通讯时间和地点，所以除了能够了解用户间的社交状况（这也是其中的一个研究方向），也能够在一定的精度范围内重构每个用户的流动轨迹，这为大规模研究人类流动规律提供了一个绝佳的实验平台。我刚加入中心那会，手机数据的研究主要是由Marta　Gonzalez这位来自于委内瑞拉的女孩负责。几个月后，她就在《自然》上发表了一篇的基于手机追踪而做的人类流动性的初步研究报告，确认了手机用户的流动性也存在列维飞行的特征，这成为当时那期《自然》的封面故事［2］。接下来的一年多时间，我开始接手这方面的工作，而Marta也顺利地找到她的下一个职位，成为了仅仅一河之隔的麻省理工学院环境和土木工程系的助理教授。

大约到了2010年初，以色列化学家Tal　Koren加入研究中心开始着手比较手机数据和Brockmann模型的异同。Tal之前的导师Joseph　Klafter是随机游走领域的专家，专精于列维飞行模型的研究，以至于同僚们都私下帮他改名，戏称他为Lévy　Klafter。Tal发现了手机用户跟纸币的流动性存在着一些显著差异，包括超慢扩散和非均匀性分布，都没法用过去的模型解释。差不多同时期，我刚刚完成一篇关于人类流动可预测性的文章［3］，于是空出手来开始针对这个问题跟Tal进行了密切交流和合作。在此之后的几个月，经过几番深入分析、计算和电脑拟合，我们发现如果把网络科学上的偏好连接引入到随机游走，就有可能解释手机数据所观察到的异常行为。这个发现导致了一个全新的随机游走模型，在相当程度上跟手机数据所记录的用户流动相符合，并发表于同年某期的《自然·物理》上［4］。关于该模型的实现细节在本书的相关章节有具

体介绍，这里只是记录了一些科学发现背后的小故事。

从这些点滴里面多少可以看到，大数据时代科学模式的转变：我们开始更多的依赖数据来找寻背后的数学规律，而不是通过几条先验假设出发试图拟合现实。这多少让人回想起约翰内斯·开普勒从他老师家第谷·布拉赫继承海量的天文观测数据，最终总结出了著名的行星运动三定律，从而开启了之后牛顿所引领的辉煌物理时代。而如今海量数据的累积，正在贯穿于社会经济文化的各个层面，这未尝不是另一个新时代变革的先兆。本书汇集了十几位来自不同领域的专家对于城市大数据所衍生的最新科研进展的报告，对包括人类流动在内的相关问题进行了不同角度的深入探讨，可以看作一个沿着大数据革命前进方向上的有益尝试。在可以预计的未来，这些尝试将逐步成为可靠的城市规划指导，在顾及宏观层面可持续稳定发展的前提下，也将为城市中的每位成员保留更多个性化的可能性和人文关怀。

2014年12月于美国迈阿密

参考文献

1．Brockmann, D., Hufnagel, L., & Geisel, T. (2006). The scaling laws of human travel. Nature, 439(7075), 462-465.

2．Gonzalez, M. C., Hidalgo, C. A., & Barabási, A. L. (2008). Understanding individual human mobility patterns. Nature, 453(7196), 779-782.

3．Song, C., Qu, Z., Blumm, N., & Barabási, A. L. (2010). Limits of predictability in human mobility. Science, 327(5968), 1018-1021.

4．Song, C., Koren, T., Wang, P., & Barabási, A. L. (2010). Modelling the scaling properties of human mobility. Nature Physics, 6(10), 818-823.

导论

作者　迪特马尔·奥夫胡贝尔　卡洛·拉蒂

译者　党霄羽

“对城市的研究，作为一种专业，在城市化经过几十年的持续发展而无处不在的时候，却消失了。当我们找到方法解释这种悖论之后，城市环境在全球范围内将获得毋庸置疑的‘胜利’。”列姆·科尔哈泽（Rem Kohlhaas，1995）

科尔哈泽在1995年撰写的文章《城市研究怎么了》认为城市化研究是一个失败的专业。尽管全球城市化甚嚣尘上，这一专业却不能改变城市的物理、社会与经济现实，甚至连一丝影响都没有。在90年代，主要的挑战是数量上的，而在20年后的今天，情况变得更加暧昧。经过50多年的郊区化对城市中心的分散，如今发达国家的城市已经走上两条截然不同的道路：要么能够在全球经济的游戏中凯旋，拥有前所未见的巨大权力；要么被工业化的限制以及人口结构的变化掏空而日益萎缩（Ryan，2012）。与此同时，在发展中国家的新兴经济体中，城市化的速度持续降低。预计到2050年，在全球范围内有67％的人口将生活在城市。（联合国2012, 2）。

城市化研究这个专业仍有许多未竟之业。无论城市以上述何种模式发展，基础设施建设、住房、社会经济发展都会带来新的挑战。但是应对这些挑战的，不仅仅是规划师、政策制定专家以及经济学家。新玩家也会登场，给这个学科送来新的研究方法。也许数据密集型领域的发展最为重要。“大数据”这一术语指的是产生于社会技术系统的大量可以利用的机读数据信息。人类和城市研究的关系变得越来越紧密：手机网络、信用卡系统、社交媒体平台，这或许是出于人们的自主选择，或许是不得已而为之。由于这些系统所产生的“数字化排放”（Digital exhaust）和我们的日常生活紧密相关，这种数字化排放已经成为一种几乎零成本的、观察社会进程和社会交往的宝贵资源。然而，由于这些数

据集数量巨大，人脑无力针对这些数据研究进行创造或架构，因此需要用新的方法来分析这些数据。

在过去10年里，随着利用这种数据资源变得越来越容易，社会科学领域发生了量化转向，这一流派被称为计算机社会科学（Lazer等人，2009）。研究者能够把计算机领域的新方法与大数据结合起来，从而研究诸如环境感知、情绪或者社会关系这样的课题，这些课题以前只能进行定性研究。计算机社会科学让科学社会学家、物理学家、数学家和计算机科学家团结在一起，这些研究者们最近发现，以城市为主题可以进行大量有趣的研究。尤其是新兴的人际关系科学，研究复杂的人际关系，在城市研究领域内产生了大量文献（Börner, Sanyal和Vespignani 2007）。人际关系科学意味着从空间视角转移到拓扑学视角来研究城市数据，关注大规模的人群、地点与机构之间的关系。在这个意义上，人际关系科学把曼纽尔·卡斯特利斯（Manuel Castells）等人倡导的城市研究变成了现实（Castells，1996）。人际关系科学通过把城市抽象为交互的社会网络空间，能够帮助我们揭示几乎所有城市系统之间的结构共性，使得研究者能够描述和预测城市如何变化，如何渐渐发展起来（Batty，2013）。在工程学方面，城市信息科学意味着利用传感器网络的城市设施（Foth，2009），包括诸如GPS到各种日常设备，这些集成技术几乎无处不在，使得我们的城市环境都能实时呈现出来。智能城市即是一个研究型学科又是一个工程学的学科，系统理论家们以及诸如IBM、西门子、思科这样的公司正在推进这一学科。智能城市这一概念向我们许诺城市管理将会通过更具适应性的基础设施建设来改善城市管理。这些基础设施能够搜集关于自身状态的信息，并且根据整个系统的状态加以调整。最后，或许也是最根本的，普通公民在管理城市方面起到的作用发生了重大改变。社交媒体的崛起，使得新的形式与社会行为加入进来。公民们自愿承担起在城市与城市基础设施的监督、管理与治理的任务，其重要性与日俱增。埃里克·保罗（Eric Paulos）把这称之

为“专家级的业余人员”的崛起。（Kuznetsov和Paulos，2010）。

由于这些发展，我们相信，有必要重新审视城市规划与研究，同时也要避免“大数据傲慢”（Big Data Hubris，Lazer，2014）的陷阱，并且关注这些新的研究方法如何给城市带来新的理解。

尽管我们讨论的这些方法产生时间不长，这些方法并非没有渊源。城市规划的历史上，有许多在新技术的激发下产生范式转换的先例。

直到20世纪60年代，控制论，关于自动反馈系统的科学已经开始在城市研究领域留下了烙印，结果好坏参半。从积极的一面来看，控制论模式为城市系统研究带来了新视角。这种模式关注动态、反馈以及系统进程，为暂时性和瞬时性带来了新的关注点，也因而在很多方面都与高度现代主义的规划理论相冲突，后者在功能上有严格的区分。但是控制模式也带来一些灾难性的失败，比如在20世纪70年代由兰德公司（RAND corporation）重新设计的纽约消防系统，造成贫穷的地区得不到足够的服务，因而导致了火灾频发、社会动荡（Flood，2010）。阿连德总统任期内旨在控制智利国民经济的Cybersyn项目空有宏大的目标，然而其问题之复杂、手段之贫乏几乎是不言自明的。

控制论明显未能抓住城市系统的社会政治维度，除此之外，它在其他方面也面临挑战：它未能借用优秀的设计理论。它容易用于模拟适应性的、复杂的动态系统，但是它没有给未来的系统提供更多的方案。安德鲁·皮克林（Andrew Pickering）认为，控制论是一种行为，而不是一种表现：它针对现状暗中做出调整，但是却不会在现有状态或者是希望的状态下对世界加以抽象的表现（2010, 19）。然而，数据为城市规划与设计开疆拓土，却是一个古老的传统，可追溯到18世纪，詹巴蒂斯塔·诺利（Giambattista Nolli）在罗马或者伊尔德方索·塞尔达（Ildefons Cerdà）在巴塞罗那所做的尝试。在人工科学方面，赫伯特·西蒙（Herbert Simon）需要一门严格的设计科学，“一门有着强大的智能、分析性的、部分形式化部分经验性的、关于设计过程的可传授的学

科。”（Simon 1996, 113）。这种设计科学有两个作用，首先对已有的设计的表现进行评估，其次是引导识别不同的场景。设计的作用是达成物理对象的“内部”世界与其目标和功能所在的“外部”世界之间的和解。“自然科学关心的是事情究竟怎样［……］而设计关心的是事情应该怎样。”（Simon 1996, 114）

关于本书

这本书要介绍一些模型，这些模型可以帮助我们捕捉那些关于理解和改进城市的各种现象。尽管大数据能够提升我们对城市体系的理解，但迄今为止，却很少有人思考设计的结果究竟如何。这本书展示了在麻省理工学院可感城市实验室（Senseable City Lab）所做的研究议程的一个剖视图。该实验室是一个跨学科研究机构，它位于城市研究和规划系，主要研究实时技术如何帮助我们更好的了解我们所生活的城市，并设想这些技术如何改善城市的可能。对本书独立贡献的作者都是（或曾经是）感知城市实验室的研究者，及经常合作者。为本书供稿的个人作者都是现在或曾经在可感实验室工作的研究人员，或者是实验室的长期合作伙伴。这些作者的学科背景各不相同，他们来自不同的专业，包括建筑与城市规划、社会学、政治学、数学、计算机科学、物理学和视觉设计等。

这本书里的文章讨论了城市数据的产生、表现及分析，并最终说明这些数据和城市规划与设计的关联。这本书相应地分为三个部分。第一部分着眼于案例研究，主要内容是城市数据的起源、构成、迭代的讨论，包括其内在的间隙和偏见。第二部分着眼于这些数据的表现形式，是利用视觉呈现还是利用数学模型。最终第三部分着重讨论城市数据对城市设计的启示。

在第一篇论文中，法比安·吉拉尔丹（Fabien Girardin）解释了数据足迹（digital footprints）的概念，数据足迹即人们使用数字服务后留下的数据。吉拉尔丹对被动的数据足迹和主动的数据足迹进行了区分，前者

是在用户无意识的情况下产生的，而后者是用户故意创造并分享的。他解释了如何用照片分享网站产生的用户数据调查旅行者的旅行行为。人们自愿上传到这些网站的照片越来越多，而这些照片中包含了明确的地理位置和即时信息（如位置标签），这就使得图片分享网站得以了解不同的群体如何出行，他们的兴趣和价值导向何在。迈克尔·赛尔（Michael Szell）和本内迪克特·格罗布（Benedikt Groß）侧重出于问责目的收集公共数据集的可能性。他们通过信息自由请求，从城市政府获得纽约市1.7亿辆出租车在一年中产生的数据集。他们对系统内大量数据进行识别，探索一种替代性的、可分享的城市出租车系统方案。在第三篇论文中，托尼·万基（Tony Vanky）（译注：即后文的安东尼·万基）考察了人们对城市数据的信任关系，这种关系有时比较模糊。目前，衡量实时城市数据的相关性——城市数据如何影响城市基础设施——的尝试还很少。范基以新加坡为例描述了衡量上述相关性的手段：是否使用并理解了实时城市数据，以及这些信息如何在个人层面影响了市民的空间决策。

大卫·李（David Lee）探讨了参与式感知（Participatory Sensing）（伯克等人，2006）数据的搜集方法。这一方法使用各种定位技术，积极地把志愿者吸引到目标数据的搜集活动之中。以垃圾跟踪项目（Trash Track project）为例，李探讨了参与这样一个项目如何反过来改变了参与者自身的行为和看法。例如，协助调查垃圾去向，是否会改变参与者对垃圾管理与回收的态度。最终，弗朗西斯科·罗哈斯（Francisca Rojas）通过对手机数据的整合绘制出纽约市的人文地图，在效度和量度方面把这些数据集与官方统计信息进行了对比。通过对电信数据的分析，她不仅绘制出了在纽约的经济中心进行全球活动的地图，还指出移民与外来工人与祖国保持频繁联系这一事实。

本书的第二部分的主题是可视化与建模，第一篇文章的作者是克里斯蒂安·克洛克（Kristian Kloeckl），这篇文章记录了新加坡实时数据平

台的发展，这一平台为搜集、整合与分配来自各种城市网络的数据流提供了便利。想知道数据的可视化如何为不同领域的利益相关者打开机遇之门，为各种城市理论的交叉提供了便利，这些“样板城市”就是具体的例子。可视化专家佩德罗·克鲁茨（Pedro Cruz）和彭索尔·马沙多（Penousal Machado）在他们的文章中表现了城市数据领域内隐喻和比喻的使用。他们侧重于关注一个地理空间可视化方面的两难问题——城市系统的空间特性与抽象的数据特性之间存在龃龉。作者探索了底层数据与表象策略之间的关系，在可视化为信息提供了“相片”还是“漫像”之间进行了区分。网络科学家菲利普·赫费尔（Philipp Hövel），菲利波·西米尼（Filippo Simini），宋朝鸣以及艾伯特－拉斯洛·鲍劳巴希（Albert-László Barabási）关注的是基于电信数据集对人类的流动行为进行观察、构型与预测。这些数据集来自移动电话网络提供商，包含了手机持有人的空间移动的隐含信息，使我们能够回答这样的问题：我们的日常行为在多大程度上是可预测的？作者通过数学分析，发现人们在出行方式上存在令人惊讶的规律，并提供了一个数学模型来描述人们的流动行为。基于沙特阿拉伯的利雅得的人群流动的案例，凯尔·格雷科（Kael Greco）对城市数据的表现策略进行了探索。这篇文章探索了城市数据的复杂性与细微差别，本文的两个角度之间相互关联而又相互冲突：其一，利用空间数据来发展观察城市的新模式；其二，利用城市的结构和组成，来探索观察和理解城市数据的新方法。

第三部分开篇是城市规划师安德烈斯·塞文随克（Andres Sevtsuk）的文章，将传统的城市设计的表现方式与以把城市作为网络加以表现的方式做了比较。塞文随克是新兴的认知学科的研究者，他展示了诸如“途中”或是“到达”这样的城市道路网络的结构措施，如何为解释城市系统内的某个地点的吸引力提供了有力的解释方法。马库斯·施莱费尔（Markus Schläpfer）的文章也在尺度研究（Scaling Studies）的领域内，对城市结构与居民和游客的出行行为的多中心问题进行了研究。施

莱费尔利用了电信数据集，调查了新加坡、里斯本和波士顿数以千计的人们出行的目的地和频率。物理学家、新兴的“尺度研究”的先锋路易斯·贝当古（Luis Bettencourt）从多个方面解释了城市系统的总体规模决定了一系列的城市特征和措施，诸如每人平均和多少人联系，人均的经济产出与创新，以及犯罪的普遍性。他的数学理论描述了城市在增大的过程中如何变化，这些变化又如何影响了城市居民。斯坦尼斯拉夫·索博勒夫斯基（Stanislav Sobolevsky）的文章关注的是人类的交流行为如何给区域和疆界带来了影响。作者利用英国、法国和比利时的例子，表明如何用手机数据，绘制地理空间的图片以表现曼纽尔·卡斯特利斯（Manuel Castell）的流动空间（Space of Flows）（Castells，1996）的概念。

这些方法能够带来城市设计的新实践吗？科尔哈泽总结道：“如果有‘新城市主义’，它不会建基于秩序和全知的幻想之上，而将会成为不确定性的舞台。它将不再对或多或少意义上的永久的对象进行管理，而是关注潜在的领域［……］它将不再沉迷于城市，而是致力于无尽的集约化与多元化、捷径与再分配的基础设施的管理——重建心理空间。”

数据驱动的计算机模型是处理与心理空间有关的不确定性、培养潜力、抓住主体以及无形特征的有效办法。数据让我们能够以前所未有的细节，建立高度动态化的城市、社会生活、基础设施网络的模型。

隐私与监视

“可以想象，如其他所有人一样，我们的组织多年以来为你能够想象的最伟大的档案中心做着准备工作，我们的档案将汇集一切，包括每个人、每个动物、所有事物，并编写目录。——卡尔维诺”（1995）

隐私和数字监控仍然是与数字系统的技术本质有关的核心问题。一般情况下，当技术产生意外的时候，隐私问题就会出现。2010年，据披露两大移动操作系统的巨头——苹果和谷歌搜集并存储其用户的位置信息。这样做的直接目的并不是监视或者目标营销：这两个操作系统都严

重依赖Wifi热点与信号塔来确定用户的位置。由于私人Wifi的热点位置很难获得，无法频繁变化，两个系统都依赖用户的手机通过自动定位每个用户经过的热点这种无形的服务获取这些信息。造成隐私问题并不是从这个服务的初衷，却是其导致的结果——产生了一个包含了每个用户行为的详细信息的巨大的动态数据库。

在早期的数字媒体，隐私问题主要和数据的产生有关——谁有权力收集数据、谁应该获得这些数据。近几十年，这场讨论转向对控制的讨论。这意味着用户应该拥有、并能够控制和自己生活有关的数据，包括有权利把自己的数据用来换取金钱或者服务。这种控制最简单的形式可以通过登入与登出来实现。关于大数据时代如何保护隐私——什么有用、什么没用，如何改进隐私的保护机制，对数字痕迹的研究也会产生一些洞见。

个人隐私问题与政府透明度问题密切相关。保护隐私需要保证公司和政府的确落实了一些规则，这些可以通过严格的政府数据公开资源得以解决。常见的需求是，政府的完全透明化，公民隐私得到完全的保护。然而在现实中，这两个领域不能完全分开，公民个人与政府行为在许多方面存在交集。

然而，上面描述的这些方法用来反抗“深层的国家势力”这样政府保密的领域是有效的。在现行美国宪法中，可以发现大数据的两种表现——民主与威胁。这个国家实现了最古老、最发达的信息自由的法律，提供了公众获取政府档案的强制机制。同时，政府通过容许在公共责任机制之外存在秘密运作的庞大的代理机构，绕开了这些法律。想通过正式与非正式的手段，把这些问题置于民主管理的范围内，就需要用高水平的数据研究来表述的公共话语，这种研究和技术与数字数据领域的知识存在区别。最后，还有个人隐私的历史与文化维度。纯文本可能是最持久、最精确的人类思想与行为的数字化表达，需要的数据存储空间却不大。由于缺乏自然的截止日期，关于数据生命周期的问题产生了——

个人数据，包括从青年时代产生的那些难堪的信息，是不是应该在某个时间被删除，或者不透露出来呢？在更大范围内，在我们这个数据记录了绝大部分信息的世界，有多少文化上的相关信息是应该被删除的呢？

卡尔维诺讲的那个试图保存全世界的记忆的例子证明，“总档案”对于所有的人类机构来说，都是不可触碰的禁地，在这个故事中最终走上了绝路。

然而，仔细思考一下就会发现，把所有数据集合起来的理想只是神话。数据的分布是高度不均的，最难获取的数据往往是我们最需要的数据。例如，在数字鸿沟的另一边的那些内陆地区。这本书中的文章说明我们的数字生活所产生的轨迹对于理解我们的城市、文化与社会有巨大的价值。可以从政府、研究机构、基础设施、公民自愿提供的数据几个渠道获得的公共数据集，我们把这些看做是公共话语的基础，并且将最终成为公共空间的一部分。

参考文献

1．Batty, Michael. 2013. The New Science of Cities. MIT Press.

2．Börner, Katy, Soma Sanyal, and Alessandro Vespignani. 2007. "Network Science," Annual Review of Information Science and Technology 41, no. 1: 537–607.

3．Burke, Jeff, Deborah Estrin, Mark Hansen, Andrew Parker, Nithya Ramanathan, Sasank Reddy, and Mani Srivastava. 2006. "Participatory Sensing," in World Sensor Web Workshop, 1–5.

4．Calvino, Italo. 1995. "World Memory," in idem, Numbers in the Dark and Other Stories, translated by Tim Parks. New York: Vintage, 135–136.

5．Castells, Manuel. 1996. The Rise of the Network Society. Chichester: Wiley-Blackwell.

6．Flood, Joe. 2010. The Fires: How a Computer Formula, Big Ideas, and the Best of Intentions Burned Down New York City – and Determined

the Future of Cities. New York: Penguin.

7. Foth, Marcus, ed. 2009. Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City. Hershey, PA: Information Science Reference.

8. Koolhaas, Rem. 1995. "Whatever Happened to Urbanism?" Design Quarterly no. 164: 28–31.

9. Kuznetsov, Stacey, and Eric Paulos. 2010. "Rise of the Expert Amateur: DIY Projects, Communities, and Cultures," in Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, edited by Ann Blandford et. al. New York: Association for Computing Machinery, 295–304.

10. Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. "The Parable of Google Flu: Traps in Big Data Analysis," Science 343, no. 6176: 1203–1205.

11. Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, et al. 2009. "Computational Social Science," Science 323, no. 5915: 721–723.

12. Pickering, Andrew. 2010. The Cybernetic Brain: Sketches of Another Future. Chicago: University of Chicago Press.

13. Ryan, Brent D. 2012. Design after Decline: How America Rebuilds Shrinking Cities. Philadelphia: University of Pennsylvania Press.

14. Simon, Herbert A. 1996. The Sciences of the Artificial, 3rd ed. Cambridge, MA: The MIT Press.

15. United Nations. 2012. "World Urbanization Prospects: The 2011 Revision". ESA/P/WP/224. New York: United Nations Publications.

第一章　数据——来源与收集

获取寰球之眼

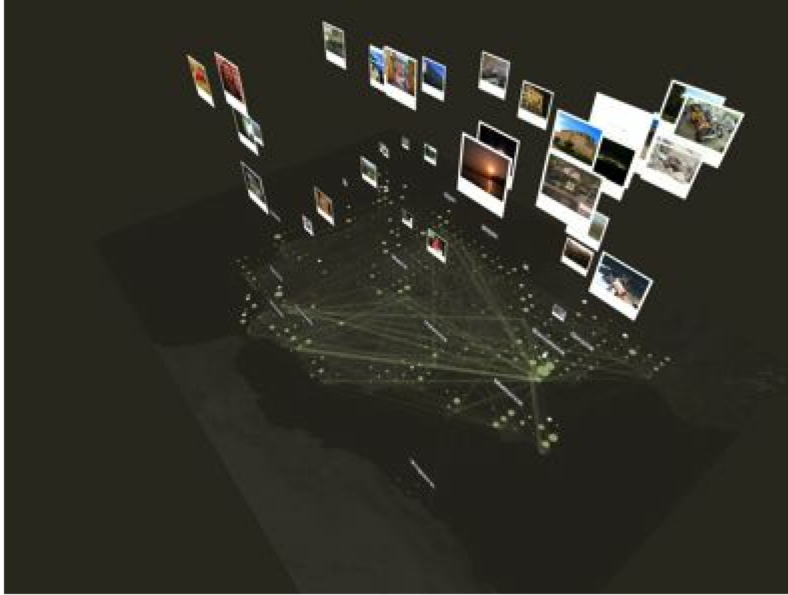
作者　法比安·吉拉尔丹（Fabien Girardin）

译者　常莹

导论

到达某一个城市的游客会通过很多种方式自愿或非自愿地留下电子足迹：在开始旅行之前，游客会查阅电子地图或者浏览网站，并生成一些日志记录；在旅行过程中，只要他们使用移动电话，无线网络就会记录下他们的行程；旅行归来，他们也许还会在网上发表一些评论和照片。一般而言，我们认为足迹有主动和被动两种。在文献中，主动足迹也被称为自愿生成的位置信息（Harvey，2013）。被动足迹是人们在使用一些能够记录位置日志信息的基础设施（比如使用移动通信网络）时留下的，而主动足迹则是用户自己在照片、短信息和一些传感设备上主动分享出来的位置数据。

“寰球之眼”项目（参见图1）研究主动足迹，它能告诉我们人们是如何游览以及感受城市的。特别要指出的是，我们使用了照片分享平台Flickr提供的应用程序接口（Application Programming Interface，API），来取得公开可用的照片。在2009年2月，Flickr共计发布了2亿张带有位置参考信息的照片（数据库中的照片总量超过30亿）。对于人们之间通过使用网络和移动设备进行互动而生成的公开可用数据而言，这一数量达到了前所未有的量级。我们收集这些“寰球之眼”，来研究游客在各个地点之间的流动，我们的研究囊括了意大利的佛罗伦萨省（吉拉尔丁等人，2008）和罗马省（吉拉尔丁等人，2008），以及纽约（吉拉尔丁等人，2009）。



【图1】“寰球之眼”项目2008年在巴塞罗那设计博物馆展出时的屏幕截图。可视化的展现方式揭示了游客到达以及移动的变化过程。大量照片汇集在一起反映出广泛的游客行为特征，它告诉我们游客在哪里、从哪里来，并且通过游客抓拍和分享的照片告诉我们他们对什么有兴趣。

用户每次将一张照片与一个实际地点关联在一起，Flickr就会为这张照片指定经纬度数值，并且根据缩放比例给出它在一张地图上的精确位置。我们认为，用户生成的内容与被动足迹不同，它为研究人群的流动提供了独特的视角。其实，一个人拍摄和遴选照片、将之上传至网络共享平台并标示地理位置的行为，比对人类的时空行为感兴趣的研究者在过去能够接触到的任何GPS日志都更加有价值。人们在他们的照片上附加了“意识的权重”，表现出了非常真实的丰富性；分析结果显示，Flickr的用户非常清晰地表现出一种倾向，他们更加乐于分享旅行中好的一面、对其间不尽人意的部分则避而不谈。“我在那里”这一类的数据，为我们提供了一种带有主观性的、反映人与空间地点之间相互关联的信息（Dourish, 2006）。

我们的工作表明，对这些数据的分析，可以告诉我们在各个时间点都有什么人到达了城市的哪个部分，这让提供个性化的服务（或者是广

告宣传）、规划博物馆的开放时间、重新分配基础设施服务以及评估某一城市政策的效果成为可能。游客自身能够感知他们所处城市当前的样子，并且会相应地采取不同的应对策略，这为我们提供了一个非常不同的视角。

从人们在现实中行为的动态地图的视角来观察问题

用户产生的内容非常容易获得，同时成本低廉，现在它对任何一个可以从深入了解大规模人群行为中受益的行业都提出了挑战。事实上，仅仅是在几年前，人们才刚刚开始在一些条件限制下讨论绘制全自动动态时空图像的可能性，这促成了人类行为数据的融会贯通以及现在这种新颖的分析方式的产生。例如佐克（Zook）等人在2004年所做的展望。

我们将许多个人的行为在城市和地区的层级上整合在一起，这些可视化的图形，可能是地理学家首次得到真实人类行为的实时地图。人们可以将它们想象成是21世纪人类社会进程的“气象图”。

主动足迹的出现意味着我们不会再丢失已经过去的瞬间；在某种程度来说，我们有了一种可以重现城市旧貌及其发展历程的新方法。这种重现城市历史的可能性，正好与近来市政当局及城市规划者对于“大数据”的兴趣相互呼应。例如，由于游客停留在旅游地期间极少留下确切的行迹，因此我们通常很难对旅游业进行量化。在寰球之眼项目中，我们得以通过对用户生成的内容进行分析以及将它们标绘在地图上，来衡量休闲型城市及其景点的吸引力。反过来说，这也揭示出了那些尚未被照片记录下来、尚未被大群游客填满的区域。照片被集合起来，反映了旅游活动的集中程度，游客在旅途中拍摄并分享出来的照片，表明了他们正在何处、来自哪里以及对什么感兴趣。

为了探索这个领域，我们进行了一系列的工作，首先就是收集数字足迹。我们使用了Flickr的API，来追溯照片的坐标及其精确位置、以及拍摄于何时，并且对于拍照人的身份进行了模糊化处理。鉴于我们对于游客的行为尤为关注，因此我们的分析平台根据拍照人在城市停留的时

间长度，将他们分为当地人与游客两组。我们对罗马进行了一项为期3年的研究，收集了6,019位不同用户上传的144,501张标示过地理位置信息的照片。我们收集到这些数据，提取了其中的时空属性，例如季节效应、游览习惯、空间分布、游客的重要流向（例如交通需求线）以及一个城市中游客感兴趣的主要地点。

空间和时间上的表现

我们将覆盖全部研究区域的数据存储在一个矩阵中，来描绘游客的空间分布。矩阵的每个元素包含照片数量和拍照人数量。在对罗马的研究中，对于游览行为的分析很快就揭示给我们游客最感兴趣的区域，例如大斗兽场和共和国广场附近的火车站（吉拉尔丁等人，2008）。时间属性也为我们进一步地呈现出游客在各个感兴趣的地点的不同行为模式。我们可以进一步猜测，到罗马大斗兽场的观光客（比如摄影师）更多在周末到来，而临近的火车站则为在工作日到此旅行的游客提供了交通工具。

交通线路需求

这项关于数字足迹的研究也让我们得以了解隐含在人们在一个城市或者地区旅游时的行走路径中的数字“公交线路需求”。基于照片的时间戳和位置信息，我们的分析平台将照片按照时间先后顺序排列起来，以重现拍照人的移动路线。更确切地时候，我们首先通过对数据进行空间聚类来找到最活跃的区域。接下来，我们将所有的个人路线整合起来，捕捉游客在游览顺序上的偏好，作为公交线路需求。我们检查每一项用户行为（例如拍照）发生的地点是否属于某一个类别，如果它属于一个类别，我们就将这个地点加入到拍照人所生成的路径中。在这个分析过程中，我们生成了许多有监督的图表以更好地支持定量分析，这让我们知道不同季节里被访问的地点数量、最经常被访问和拍照的兴趣点、拍照人在何处开始以及结束他们的旅行。

兴趣点

之前的研究已经证明了，网上能够获得的、标注过空间和时间信息的资料，可以被用来发现与语义信息有关的“地点”和“事件”（Rattenbury等人，2007）。通过类似的方法，我们可以通过与用户生成的照片有关的标签，来了解人们对于他们所处环境的看法、他们对于城市空间的观点的语意。例如，“废墟”是最常被用来形容罗马的照片的词。我们画出了2,866张照片的标签的分布，找到了罗马城里最古老、最“残破”的部分：大斗兽场和古罗马广场。在对2008年纽约人造瀑布项目[1]的经济效应进行的研究中，我们使用这样的语意信息来找到拍照行为发生的主要区域。

案例研究：衡量一个事件的影响

我们对2008年发生在纽约城的人造瀑布公开展览进行了一项研究，来进一步探索能够被用来定义数字足迹的、清晰的指标，以衡量城市吸引力的演变。市政当局的目标，是要比较各个观景有利地点[2]的吸引力和受欢迎程度的变化情况。因此，我们衡量和当地居民与游客的空间分布，并且通过比较数字足迹表现的变化，来证明纽约人造瀑布项目对于水岸区域的吸引力有促进作用。最后，两个主要的结果提升了纽约对于这一事件的效果评估：拍照者到达情况所表达出来的地点吸引力变化情况，人群聚集情况所表现出来的地点受欢迎程度的变化情况：

根据拍照者的出现而得到的地点吸引力的变化情况

我们基于拍照者在各地点出现的相对情况，我们的研究分析了衡量吸引力的指标的变化情况，这些指标是根据2006年夏天拍照者在各地出现的情况而得到的。研究显示，与纽约城的其他地区（如时代广场和中央公园）相比，水岸地区的吸引力在2007年夏天和2008年夏天分别上升了8.2％和20.7％，这表明了人造瀑布的展出所带来的潜在影响。

基于人流集中度的受欢迎程度的演进

人们对一个区域的兴趣的集中程度决定了它在大众的拍照者人群中受欢迎的程度。我们的地点排名（PlaceRank）指标显示，相较于2006

年，2007年一些受欢迎的地点的人流集中度下降了15％，而同时另外一些地点的人流集中度则上升了10％。而与2007年相比，2008年这些地点的人流集中度则上升了56％，同期其他区域该指标下降了30％。在2008年，在这些受欢迎的地点像其他游客感兴趣的地点一样，出现了人流集中的情况，这意味着它们与那片城区的其他兴趣点一样，都出现在了游客的旅游路线上。

这项研究表明，数字足迹的出现为我们带来了一个详细评估空间使用情况、事件影响力以及建筑环境的机会。这一方法不但能够让我们做出更好的城市设计决策和城市管理决策，而且让市政当局能够取得及时的信息来了解民众对空间的使用情况以及评估城市结构改造的影响。事实上，综合我们对纽约人造瀑布公共艺术项目的经济效应的各种研究结果，可以发现，我们在分析中所使用的那些指标为传统的研究方法提供了有益的补充。

讨论

无处不在的科学技术不但为我们的日常生活带来了便利，同时也为我们对于时间与空间的研究提供了方法。通过用户生成的内容，我们可以更好地了解城市环境下人群的流动性，以下我们要对它的一些意义做重点说明。

技术意义

除了我们对于城市吸引力指标的研究，其他一些研究小组使用了现实挖掘[3]的方法来发现具体的城市动力学特性。使用这类方法时所面临的最大挑战在于，对于数据的边界和偏差达到清晰的认知。我们行为并不表示认可，我们不能将两者混为一谈。举例来说，在我们对于纽约人造瀑布的研究中，我们使用了数字足迹的密度作为衡量城市吸引力的指标，这是项目的一个局限所在。因此在未来的研究中，由可靠的技术所提供的地面实况信息将用来对数据进行校准。

此外，有些分析报告建议将有地理坐标的照片与有地址标注的照片

拆分为不同的部分。拆分可以根据社会文化、国籍、游客类型（由游客在城市停留时间长短以及他们对城市的熟悉程度决定）、游客对科技知识以及空间导航技能的掌握水平、游客要达成的目标以及他们所访问地区的环境情况进行。此外，我们也应当考虑游客在旅程中有多大的可能性可以使用他们的移动设备来生成数据。对于这类问题的解答，让我们可以更好地界定数据的意义，并且进一步探索它们在科学以及城市研究方面的潜在应用。

方法论方面的意义

重现城市的能力表明，研究者有机会创造出一些新颖的方法来研究城市环境。当然，这里有一个非常重要的假设，那就是认为世界是由数据碎片组成的，这些数据可以被加工成为一些信息，而这些信息本来就会为人们带来某种价值。这将引领我们走向数据驱动的城市生活，如果城市生活可以被数据驱动的话。

事实上，了解一个城市远不止是记录设备状态和事件而已。因此，我们不能将形式新颖的绘图技术的发展，与我们之前提到的一些与技术发展水平有关的关键事项混为一谈。现阶段我们仍然努力想要弄清楚：1）数据能够向我们揭示哪一部分的现实世界？2）我们可以用数据来做些什么？3）为了获取信息，我们应当如何与人们沟通（这还远远不能被称为是“智能的”）。

即使是考虑到这些问题，应用我们的研究方法仍然很有希望能够取得关于人群如何到达以及出现在某地的知识，而且通过一些特殊的技术我们也将创造出一种“以人口／数据为基础的城市化”的方法。这种方法将被用于以下方面：

1．定性的分析启发了定量的研究：这种方法最初聚焦在研究人们和他们的行为，我们并没有做出与数据的处理和计算相关的假设，这是研究的一个缺陷。这个定性的研究角度提示我们接下来要进行定量的分析来取得更多有关特定人群的行为以及模式的经验知识。在这个领域，有

一些方法能够满足这个方向的需求。例如，William等人（2008）认为，在研究人们在城市中的体验时，对于个体在城市中感受的研究，比将某种特定的城市形态作为研究的起点，更能够帮助提升我们对于城市的了解。

2．定量的数据挖掘启发了定性研究：在上述的方法中，定量数据帮助我们发现新涌现出来的行为以及异常的行为，主要作用在于发现问题。接下来，定性的研究视角可以帮助我们来解释一定条件下所发生的现象。定性的研究方法实际上要求我们能够提出一个正确的问题，去发现特定情况下各种有意义的情形。例如，在纽约人造瀑布影响力的研究中，我们就使用了定性的方法。我们使用了横跨3年时间的数字足迹。作为对这一定量分析的补充，我们可以对被监测地区进行定性的观察，以解释地区的吸引力是如何发生变化的（例如，研究人们是否会停留更长的时间？）。

这催生了人们对于发展出一个与行为痕迹有关的、完整的知识体系的需求：既包含定性的方面（例如，与行为和谈话有关的音频与视频），也包含定量的方面（例如用户主动生成的内容）。空间使用的实际数据具有重大意义，我们可以通过这些数据来进行一种新型的“使用后评估”，在城市规划和建设的实践中我们时常会忽视这一部分（Brand, 1995）。然而，作为方法主要组成部分的那些工具、指标和解释方法，仍然有待进一步的发展。

社会意义

大量存在的地理位置相关信息已经广泛地被授权（由那些能够生成以及消费这些信息的个人和地区产生），并且可能已经过分地被授权了，人们对此习以为常，国家管理机构有可能更好地规范这些不断增长的个人和空间个性化信息。因此，我们需要努力应对随之而来的伦理问题和隐私问题。由于人们对于可追溯性的理解，他们对于寰球之眼项目中与绑定了地理信息的数字足迹有关的研究感到忧虑，这种担忧是合理

的。尤其我们的研究工作展现出了对于隐私的两种监控方式的差异，一种是自上而下的、老大哥式监控，一种是小范围内人们自下而上的、小女孩般的观察[4]；它使得我们选择不使用一种技术的决定，等同于选择退出一个社会。

事实上，这些数字足迹在当代社会是一种必然的存在，而且，如果我们希望能够享受现代生活的便利，那么它们也是必须的；我们不可能再与它隔绝开来，就像我们无法在艳阳高照的白天甩掉自己的影子（佐克等人，2004）。数据这个影子的发展是一个难以清晰辨识的过程，在这个过程中，会存在不同程度的对于个人隐私的担忧，以及在各种情景下以隐私来换取便利的自发行为。

总体而言，无处不在的地理位置相关信息为我们提供了绘制地图以及模拟人类动态迁移行为的新方法，同时，它也对我们当前对于隐私的观念提出了挑战。我们应当重视这些挑战，并且通过保持地理位置相关信息的复杂性和丰富性，来避免将这些信息整合为可供独裁者使用的统一结构。

参考文献

1．Brand, S. (1995). How Buildings Learn: What Happens After They're Built. Penguin.

2．Dourish, P. (2006). "Re-space-ing Place: "Place" and "Space" Ten Years on.," in CSCW, pp.299-308.

3．Girardin, F., Vaccari, A., Gerber, A., Biderman, A., and Ratti, C. (2009). Quantifying Urban Attractiveness from the Distribution and Density of Digital Footprints. International Joural of Spatial Data Infrastructure Research, 4, pp.175-200.

4．Girardin, F., Calabrese, F., Dal Fiore, F., Ratti, C., and Blat, J. (2008). Digital Footprinting: Uncovering Tourists with User-generated Content. IEEE Pervasive Computing, 7(4), pp.36-43.

5．Girardin, F., Dal Fiore, F., Ratti, C., and Blat, J. (2008). Leveraging Explicitly Disclosed Location Information to Understand Tourist Dynamics: A Case Study. Journal of Location-Based Services, 2(1), pp.41-54.

6．Kostakos, V., Nicolai, T., Yoneki, E., O'neill, E., Kenn, H., and Crowcroft, J. (2008). Understanding and Measuring the Urban Pervasive Infrastructure. Personal and Ubiquitous Computing.

7．Rattenbury, T., Good, N., and Naaman, M. (2007). Towards Automatic Extraction of Event and Place Semantics from Flickr tag. In SIGIR'07: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, S. 103-110, New York, NY, USA. ACM Press.

8．Ratti, C., Pulselli, R. M., Williams, S., and Frenchman, D. 2006. Mobile Landscapes: Using Location Data from Cell-phone for Urban Analysis. Environment and Planning B: Planning and Design, 33(5), pp.727-748.

9．Williams, A., Robles, E., and Dourish, P. (2008). Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City, Chapter Urbane-ing the City: Examining and Refining the Assumptions Behind Urban Informatics, Hershey, PA: Information Science Reference, IGI Global.

10．Zook, M., Dodge, M., Aoyama, Y., and Townsend, A. (2004). New Digital Geographies: Information, Communication, and Place. Geography and Technology, pp.155-176.

Hubcab——共享出租车服务的优势

作者　麦克尔·塞尔（Michael Stell），本尼迪克特·格罗布（Benedikt Groß）

译者　黄艺

运用大数据研究人类流动性

人类活动的数据在改善城市交通系统方面潜力巨大。人类的互动，比如电话通话、信用卡交易，或者是线上社交网络的使用，衍生出了大规模的数据，这些数据揭示了人类日常行为和流动模式，使其显现出高度的可预测性（Song等人，2002），并且使得我们可以其背后隐藏的统计“定律”。如今关于人类流动性的研究，正如其它关于人类行为方式的研究一样，已经不仅仅是社会学的关注点；人类流动性的话题正在更技术化的领域内吸引更多人的兴趣，比如说有计算机科学、数学或者物理背景的研究人员，来研究人类个体的流动模式，如同研究那些无生命的基本粒子一样。令人惊讶的是，像人类这样高度复杂生命体的整体统计行为在很多情况下能够用完全机械化的术语来理解和表述（Ball，2003）。

但是现在所有这些太字节（Terabytes）的关于人类流动性的数据是怎么变得如此唾手可得，而基于这些数据设计交通系统的结果又会如何呢？这些数据的即时可得性根本上是因为科技和社会的进步：移动电话在一方面变革了我们交流的方式——我们可以随时被联系到，但是另一方面我们也总在留下踪迹。电信业业者运用了强大的基础设施，在我们每次通话和发短信的时候记录离我们最近通信基站的位置。在每个新型手机和移动电子设备上装备的全球定位系统（Global Positioning System，简称GPS）技术使得对于位置的追踪更加精确。能够追踪的不仅仅是人们步行的路线，还包括人们搭乘任何城市交通系统的路线。在交通工具上安装的追踪器可以绘制出图像，以表示一个城市里数百万车

流的壮观景象（图1），让我们更好地理解人类的集体行为以及我们交通系统的瓶颈问题。



【图1】纽约市的出租车车程轨迹，截图来自hubcab。纽约市的街道被分成了40米长的部分，所有2011年整年的纽约市出租车活动都被可视化了。如果一个街道有更多上出租车的乘客就呈黄色，如果有更多下出租车的乘客就呈蓝色。街道部分的宽度和出租车活动量成正比，出租车到过几乎所有的道路。主干道更倾向于呈黄色，然而小路更倾向于呈蓝色，描绘出一个复杂城市系统美丽的图像。

在社会发展方面，让我们重点关注一下所谓的开放性数据和开放性政府的倡议。开放性数据的概念是指特定数据对于任何想使用该数据的人都应该可以自由获取，而不用受版权、专利和其他控制方式的限制。

在几个最近的案例中，一些行政和政府机构向我们证明了公开数据这一步是明智的。这些案例包括：公开城市公共交通系统的数据带动了智能手机上关于实时公车火车时间表应用的独立开发，open311的企划让居民更直接地和他们的城市互动［5］，或者是在开放数据目录（Open Access Directory）列出的各种旨在让不同领域科学家使用的数据库［6］。一脉相承地，“开放街景”（OpenStreetMap）是一个在（数据）开

放许可证允许下，在地图上绘制并发布制图数据的合作型项目[7]。在美国，现代开放性政府的理念可以追溯到1966年，美国总统林登·贝恩斯·约翰逊（Lyndon B. Johnson）签署了信息自由法（Freedom of Information Act）这份联邦法案。在此之前，国会议员John·E·莫斯（John E. Moss）在十年间发起了多次众议院听政会，一直试图让艾森豪威尔（Eisenhower）领导下的政府审议通过这项法案（Blanton，2002）。Moss的努力让得到了很多知名记者的支持，记者们希望消除官僚的阻碍和当时政府的黑箱操作。莫斯尖锐地指出，“你得削尖脑袋才能拿到一点信息”（Kennedy，1978）。信息自由法使得以前受美国政府控制未公开的文件得以公开，并且设定了强制性的公开步骤。虽然该法案在那以后被修改过很多次，然而它主要的准则在今天仍然生效，也让下文中我们的很多研究可以进行。水门事件（Watergate）和其他影响深远的丑闻使得信息自由法能够继续存在，并且该法案现在仍然是全世界许多开放信息运动的催化剂。

除了以上这些为公众利益开放更多数据这样令人满意的系统性构想，研究人员使用的数据源通常仍然依靠私人的或无计划的途径得以公开。举例说明，移动电话的信息通常是电话公司专有的数据资产——如今它依赖于研究团队的说服力，以及公司的态度：代表公众利益的信息是否可以以研究的目的公开。一个最好的例子即2003年Enron Corpus[8]的公开，它证明了意外的数据公开可以产生众多社交网络和机器学习领域的研究成果。2003年，在对于Enron公司崩盘丑闻的调查以后，Enron Corpus这个由158名Enron员工的600,000多封邮件组成的巨大数据库被公开。针对于Corpus展开的协同社交网络研究发现了原先未被记录的公司操作行为，并且可以帮助洞察在大型组织里普适群体动力学（group dynamics）的本质。

在流动性数据方面，2011年4月人们发现[9]苹果iPhone可以追踪和存储用户几个月的行动记录，而且这些记录容易被获取。这就是一个无

计划数据公开的最好例子。发现这些行动记录使得很多类似于crowdflow. net的企业兴起，致力于创造关于wifi和电话网络公开数据库［10］。

改进城市出租车系统

我们该如何用这么多的数据来改善城市生活呢？在人类流动性的话题上，显然我们感兴趣的是城市交通系统的运转。如今城市化愈演愈烈，交通流量迅速增加，这使得城市交通服务的流畅运转比以前任何时候都要重要。交通堵塞是限制城市环境的主要因素，汽车释放的尾气影响了当地的空气质量，产生了显著的经济和环境影响。城市出租车服务是城市交通最主要的方式之一，也是下面我们主要的关注对象。尽管出租车行业在城市交通中很重要，但以前学术圈对它的研究兴趣有限，并且该行业在过去抵制了系统上的变化和进步。举个例子，几十年来几任纽约市市长都曾经承诺要改进出租车系统，但是出租车行业自1900年以来本质上都是一样的（Li，2006）。然而不管是对于出租车公司和它们的客户，还是对于城市规划师们而言，提高出租车派遣系统的效率都是有重要意义的，比如说可以运用动态分配出租车的程序来减少燃料的消耗，提高服务质量和改进城市交通的可持续性。

在最近的一项研究中，我们研究和运用了一个包含了2011年纽约市（New York City）1.7亿出租车旅程中13,500辆出租车位置的数据集。纽约市的出租车业务是由纽约市出租车和加长轿车委员会（Taxi and Limousine Commission，简称TLC）监察的。TLC向出租车严格限制许可权（又称“圆牌”）的发放，这些许可证赋予了我们熟悉的黄色汽车独有的，在街上人们招手叫出租车时能够接上乘客的权利。自2008年起，纽约市所有的出租车被TLC强制要求安装一个出租车科技系统（Taxi Technology System），其中包含一个GPS追踪器和一个在后座屏幕上让乘客看的、标有实时汽车位置的地图。追踪器收集的GPS数据被直接传至TLC。

对这些GPS追踪器的安装并非毫无争议。在这之后出租车司机们时而罢工，状告TLC，并且这种情况持续了好多年。原告们的集体诉讼称安装GPS追踪器是“反宪法的，未经许可地侵犯了他们的隐私”。在这之前，TLC根据收集的数据认为有些司机对乘客过度收费，而撤销了一些“圆牌”许可证[11]。无论如何，TLC成功地实现了这项改变。现在根据信息自由法案，任何人都可以向TLC要求完整的、匿名化后的出租车位置数据。

纽约市不是唯一一个拥有出租车位置数据的城市。在新加坡－麻省理工学院研究和科技联合企业（SMART）中，新加坡最大的出租车服务商对一些特许的合作伙伴开放了从新加坡26,000辆出租车中16,000辆出租车上收集的数据，以提高新加坡市的生活质量。类似的研究还囊括了上海、旧金山或者维也纳的数据，而且预期这类数据在以后可能会越来越多。由于纽约市数据高度清晰，并有一整年度的完整数据，我们首先集中研究纽约市出租车车队的数据，然而我们的研究结果可以不失一般性地应用在任何主要城市区域的出租车车队上。

对于纽约市数据的研究立即揭示了出租车系统作为一个整体是效率极其低下的，如同我们在许多其它城市发现的那样。其效率的低下程度可以由没有搭载乘客的旅程，即空车旅程的总距离和总时长来量化。空车旅程长度的分布呈现双峰分布：第一个峰值对应的是可预期的平均空车长度，然而第二个较小峰值对应的大概是从城市中心到机场的距离。这些特殊区域，比如机场，城市中心这样的密集区域，是造成出租车供需关系不平衡的主要原因。出租车司机倾向于集中在几乎可以保证能接到客人的区域，比如说主干道或者其它像机场一样人流量大的地方。然而由于信息传播速度缓慢，实际接到客人与否的不确定性大大地提高了，尤其是在机场，某些时间段，想到要去机场的司机过多，空出租车的队伍可能会变得非常长。系统的低效率性自然而然产生了不利的影响，比如尾气排放量高出城市正常运行所需的排放量；对于道路系统的

低效率利用也使得堵车、道路瓶颈等等现象更加扩大。机场和类似地区引发的交通系统的显著问题，可能会影响在未来城市规划时，重新考虑将这些区域放在战略上更理想的位置。

以前也有一些试图改进出租车系统的尝试，大多数时候是以自发的形式。在许多亚洲城市，拼车是一个众所周知的现象，乘客可以放弃一些舒适度和行程时间，分摊乘车费用。更近一些，商业上也有联运交通的尝试：比如说Daimler公司的Moovel[12]服务就是联合包括出租车在内的几种不同交通方式，也有公司开始提供汽车分享服务，然而这些尝试总是受到人们接受度和类似上面提到的效率问题的困扰。另外，利用一些智能手机应用，按一个按键就可以很容易地发现、呼叫出租车，并且给出租车旅程付费。这些应用在全世界各个城市里的使用情况良莠不齐。然而城市出租车系统作为一个整体而言，看上去并没有展现出大幅的变化。

正是因为上面提到的出租车系统的死板和低效率，我们提出一个新的系统以取代旧的系统或者和旧的系统同步运行。因此我们运用了可获取的数据和数学模型，设计出一个更高效的新系统，能产生更少的尾气排放量，而且较旧系统而言对于顾客更加实惠。纽约市数据中的出租车旅程覆盖了城市99％以上的街道（除了比如史坦顿岛之类的区域，那里是没有出租车服务的）。每一段街道都有一些特定的目的点（出租车将要去的其它街道）和出发点（出租车从其它街道来到这里）。特别是我们发现非常多的旅程是有可能拼在一起的，从而可以减少这部分旅程。在曼哈顿的密集区域，很多街道交叉口在一年之内有200万次出租车旅程在这些街道的目的点和出发点附近开始或者结束。其中很多旅程在相似的时间开始并结束，如果乘客们愿意分享同一个出租车，这些旅程就是多余的。所以我们的方案致力于开发一个新的出租车派遣算法，基本的想法就是合并出租车旅程。

这个新的过程旨在最小化出租车服务的总花费，当然，仍然是在满

足一定舒适度的前提下，比如说乘客不会比原来需要多等候一定百分比的时间。在最简单的情况下，一辆出租车能容纳两位乘客搭载。当两位乘客在同一出发点希望到同一个目的点，并且在同样时间，出租车就可以同时接送两位乘客。

当然可以放宽一些假设，举例说明，两位乘客的起始和目的点不一定要完全吻合，或者是一个乘客的起始和目的点之间的车程比另一个乘客更长。不管怎么说，我们先通过模拟展示了现有出租车系统在一辆车两位乘客的搭乘容量下可以被优化，得以更高效地消化人们对出租车的需求。自然而然地我们也可以提高出租车载客量到任意更大的数字。如果模型、模拟和分析性调查显示更大的汽车容量更加高效，实际对于大载客量出租车的实现可能偏向于“出租车加长轿车”。每位乘客在大车中有自己的私人空间，与公车类似，但是线路是个性化的、动态化的。数学上来讲，我们的方案定义了一个旅途分享性的网络，网络节点是出租车车程，节点间的连线表示两个车程在一定规则限制下有可能组合在一起。当有两个以上的旅程被分享和组合的时候，这个网络变得更加复杂，会有“高维度”的连线。然后，我们的算法解决一个叫做最大化配对的问题，就可以算出车程合并这个问题的最优解。对于大型（网络）图的精确解在计算上是不可行的，然而有算法可以在合理的计算时间内估算最优解。在这个计算过程中，我们定义一个延长参数，表示一个乘客愿意延长的最长的旅程时长。如果选择一个短时等待，服务质量会更高，因为乘客可以等候更少的时间，然而可以被合并的旅程也更少了。延长时间的话，情况会相反。

这项研究的目的在于在不大幅降低服务质量的前提下，评估分享出租车的潜力。对于曼哈顿区域车程的初步研究结果是鼓舞人心的：在分享两个旅程的情况下，如果乘客愿意最多延长旅程5分钟，90％以上的旅程可以共享，可以减少总旅程长度以及产生的污染40％以上。在分享三个或者更多旅程的情况下，出租车分享的潜力看来更大。旅程最多延

长五分钟，换来的是花费减半，这看起来合情合理。我们注意到，在线上分配系统中实际运用旅程分享的算法似乎很简单。乘客叫了出租车以后，几乎可以马上收到他分享出租车的选项。

虽然可分享的城市出租车系统在技术细节上的问题似乎可以用算法来解决，然而说服权威机构和顾客来使用新系统这项重大的政治任务能否完成，仍然是一个开放性问题。

用于改进发展现有技术的前期投入可能会被认为太高，而且可能需要创建鼓励机制（除了分摊车费以外）帮助人们克服与陌生乘客接触的恐惧感，比如在每辆车中建立分开的、私人的隔间。

HubCab工具

在我们进行研究的同时，我们也在www.hubcab.org开发了一个叫做“Hubcab”的免费线上工具，以便于让更多受众用简单和有趣的方式来探索我们研究的现象。上文描述的纽约市出租车旅程利用这一工具得到了可视化处理，显示了1.7亿次汽车车程，并且让人们可以从一个全新的角度、互动的方式探索这个城市，向全世界的受众展示了一个智能分配出租车算法的潜力。

我们特别以一种直观的方式表现出租车旅程的数量和方向。出租车卸客和上客的过程被条理清晰地、动态地、有层次结构地表达出来，在更高聚焦的层面展示了单次卸客和上客的地点所包含的最多信息。这个方法的优点是使用者可以同时看到旅程的方向和估算的总旅程数量。我们用了openstreetmap. org的地图作为地图数据。我们开发了一个脚本来将纽约市丰富的道路网络划分成200,000个40米长的部分，以给使用者提供高分辨率的体验。这么多街道部分中，每个部分都包含了至少一个放下或者接上乘客的地点，有些部分甚至包含了上千个（在一年的时间内）。Hubcab可以放大地图，以前所未有的大图，显示单次放下和接上乘客的地点（图2）。使用者也可以在不同时间段之间切换，显示整个时间长度的数据或者不同时间段的数据，以突出在一天不同时间里，不

同的接放乘客的“热点”。周边区域的半径也可以在显示数据的同时随时改变。使用者可以动态选择所有在接放地点之间的出租车旅程。Hubcab让人们可以探索400亿种组合（200,000乘以200,000）的街道间的出租车流动，并对其进行可视化。



【图2】Hubcab工具的不同缩放程度。高度数据空间分辨度可以允许很大的变焦程度，可以显示在纽约市街道层面上单次的出租车接放乘客。

从出租车系统这个新的切入点，Hubcab这个工具对城市内部的运作提供了前所未有高清晰度的独特见解。Hubcab可以用来调查到底出租车

在何时以何种方式上客或卸客，并且可以用于确定密集接放乘客的区域。最重要的是，在城市任意两点之间车流量的可视化让市民可以感受到很多空闲的车程和出租车系统巨大的改进余地。Hubcab扩大和改变了人们对于城市空间的感受，并且有潜力改变城市居民和访问者的行为。Hubcab在社会和政治上的影响是明显的；通过使用这类工具，城市的规划可能会显著改善。这些可感的、可视的数据可以被用于更好地设计城市，在系统层面上更好地规划出城市未来的原型。在我们出租车系统的案例中，有希望看到这种改进能减少道路交通堵塞，带来更低的运行成本，以及更少污染、更加清洁的环境。

在我们的研究方法中，我们对（网络）图使用了最优化算法。除了这个逻辑上的优化方案，其它还有哪些可能解法呢？先前用移动电话数据的研究表明，特定的空间符号会对人类活动的地理环境产生重大的影响（Reades等人，2009），从而改变交通系统的运作。我们也可以考虑重新安排城市结构，改变土地使用的模式，以实现改进城市交通系统为目的的整体方案。在任何情况下，只有对现状进行评估，对取得数据进行严密的分析，才能实现高效、系统性的改进。

致谢

作者感谢合作研究者Paolo Santi和Giovanni Resta（意大利比萨的CNR数据和电信学院）、Steven Strogatz（纽约州伊萨卡的康奈尔大学数学系）、斯坦尼斯拉夫·索伯立夫斯基和卡洛·拉蒂（可感知城市实验室）、Hubcab网络开发员（47nord），以及国家科学基金会、AT&T、MIT SMART项目、MIT CCES项目、奥迪大众、BBVA、爱立信、Ferrovial、GE和所有麻省理工学院可感知城市实验室的成员对于研究的支持。

参考文献

1．Ball, P. 2003. The physical modelling of human social systems. Complexus 1, pp.190-206.

2．Blanton, T. 2002. The world's right to know. Foreign Policy 131, pp.50-58.

3．Kennedy, G. 1978. Advocates of Openness: The Freedom of Information Movement. Dissertation, University of Missouri-Columbia.

4．Li, S. 2006. Multi-attribute taxi logistics optimization. Dissertation, Massachusetts Institute of Technology.

5．Reades, J., Calabrese, F., and Ratti, C. 2009. Eigenplaces: analysing cities using the space-time structure of the mobile phone network. Environment and Planning B 36, pp.824-836.

6．Song, C., Zehui, Q., Blumm, N. and Barabási, A.-L. 2010. Limits of predictability in human mobility. Science 327, pp.1018-1021.

数据可用性与相关性：衡量新加坡实时城市信息的使用

作者　安东尼·万基（Anthony Vanky）

译者　曹海元

实时城市技术如日方升[13]。交通与运输信息，环境质量测量，甚至是城市情绪的信息（挖掘自社交媒体[14]）：所有这些信息都可以被实时监测。这些传统上用于优化和管理的系统，现在对普罗大众越来越触手可及。移动设备和个人技术可以让用户知晓那些他们每天接触到的基础设施与资源的当前情况。

通过新型个人技术获得的这些数据允许个人根据他们的当前环境来改变行为。举例来说，如果某人想要通过在特定时刻离家来错开交通高峰期，那就可以依据实时信息而不是常规习惯来规划离开时间。个人也可以依据由气象站站得到的实时信息来延长一点工作时间，避免在下雨的时候乘坐公交车。

虽然信息的容量和范围增加了，然而针对这类数据相关性进行的公共性测量还很少：它如何改变了人们与城市基础设施之间的互动。我们注意到了用户的数量，但是忽视了使用的性质和方式。这项研究的目标是描述那些测量：实时城市信息是否值得监测与使用，以及信息如何影响市民在个人程度上进行决策的进程。这里我们采用新加坡居民这一特定的人群作为我们的案例，即这项研究的主体。

借用工程学上的一个术语，这项研究关注的是数据的驱动力。测量出实时数据应用的技术，诸如API（译注：Application Programming Interface，应用程序界面）请求的记录或者是数据传输的数量，已经建立起来了。这项研究超越了数据的流动，而是致力于考察终端用户：任何由数据使用导致的感知与行为。进一步来说，当一些关于实时信息的研究专注于孤立的技术领域，例如孤立的运输问题时，我们仍然有机会

从数据的获取及数据最终的实际应用这两个方面对该研究的意义进行更广泛的理解。

对很多规划者而言，这些技术服务的意义是能够使市民个体更顺利地在城市中移动。人们可以凭借具体信息，而不是直觉或者记忆来增强他们与城市环境的互动。另外，他们能够反馈所使用数据服务的表现。越来越多的城市开始在公共交通上装备实时信息反馈。能源公司装备了所谓的智能量表，它可以让能源用户实时监测他们的使用情况。技术创新的源泉是创造更多的产品去卖给那些城市，以实现各个城市监测与测算的目的。虽然这些技术的应用已经前所未有的广泛，然而当前的分析无法测量出终端用户的体验。研究还是依赖于间接地测量，根据软件应用在网络上进行API调用的数目或者递交报告的数目来解释其普遍的应用，而不是根据终端用户，即城市居民的体验。

人们对于这些技术往往会顺理成章地假设“如果你实现了技术，用户就自然会到来”——即提供开放和实时数据的机会越多，使用的市民数量也越多[15]。这些技术的增长是毋庸置疑的。在这个10年结束时，预期会有1080亿美金投入到智能城市技术的应用中[16]。然而，在没有理解用户体验的情况下，我们不能知晓这些工具在何地产生了怎样的影响，以及这些投资是否对市民日常生活产生了实质性的作用。这些测量可以让我们通过了解到底人们想要达到多么高的效率，来更好地理解数据能够通过什么方式激励个人更高效地生活。

案例：新加坡

新加坡是一个独特的，具有潜在前瞻性的案例。文化和经济的因素使得这个国家特别适合这项研究。在这里，测试个人用户如何使用及评价可得的实时数据成为了可能。

对技术的关注在国家层面根植于新加坡与其他地区进行贸易和商业交易的历史地位中。马来亚、沙巴、沙捞越和新加坡在1963年联合组建了新国家马来西亚，而新加坡这个城市国家在1965年从马来西亚分裂了

出去，成为了一个完全独立的共和国。紧随独立而来的是英国的部队和岛上最大雇主们的离去，一种经济增长方式应运而生。这种经济增长是通过对外来直接投资的极度依靠和多国对廉价土地和劳动力的渴求实现的。这也导致了新加坡异军突起成为亚洲经济“四小龙”之一。在1980时代后半叶，新加坡经济开始从廉价生产力向强势的服务业转型，外加“高附加值”先进制造业、金融业和贸易业。为了促进转型，经济规划者们意识到了信息技术作为发展的战略助推器的重要性。

政府在1981年建立了国家计算机委员会（NCB），“驱动新加坡在信息时代通过大量运用IT技术增加经济竞争力和国民生活品质，达到脱颖而出的目的。”[17]在1986年国家计算机委员会发布了首份国家信息技术（IT）计划，目标是发展其公有和私有部分的信息技术能力，与此同时增进国家整体的经济竞争力。下一个计划，在1992年发布的《一座智慧岛的展望：IT2000报告》考察了一个更重视普及与互联的信息技术如何把新加坡打造为一个世界主要经济中心之一[18]：

“在我们的展望中，从现在起大约15年时间，新加坡，一座智慧岛，会成为最早具有国家范围先进信息基础设施国家中的一员。每个家庭、办公室、学校和工厂中的电脑都会实现互联。”[19]

比简单出口导向的经济计划更为宏大的雄心是基础设施，即新加坡整体发展所不可或缺的部分[20]。政府已经估算了在2013年公有部分在城市信息和通信技术（ICT）基础设施上的花费会达到120亿美金，比5年前提高4.8％，额外还有26亿美金作为一个更宏大的技术总体规划的一部分[21]。然而，伴随新加坡经济增长而来的是其城市基础设施所面临的持续增长的压力。更通俗的说，新加坡成长不会如此的简单。大量的技术投资流向了管理和操纵岛上市政设施的实时城市基础设施。经济和城市境况合力把新加坡变为了“智能城市”。

当政府对技术与获取数据作做出强有力的承诺后，仍旧缺乏一套方法对其目标公众群体的接纳程度和使用感受进行评估。新加坡提供了一

个良好的环境来测试实时信息的潜力；这是一群精通技术的人，从而不用担心缺少能够能投入使用的技术。因此，新加坡为评估实时信息的使用感受提供了一个得天独厚的环境，因为这里可以减少对一些传统上人们在接纳和使用科技时壁垒的担忧。

接受调查的受访者所持有的移动电话数目已经超越了受访者的人数。这个结果在新加坡并不令人惊讶，因为移动电话的持有率已经达到了每100人145.24部[22][23]，每10个人中大约有7个人可以连接上互联网[24]。这项研究也发现许多人熟悉“实时信息”的概念。当个体在这项研究中被问到他们自身对“实时信息”和“实时数据”的熟悉程度时，59.6％回答对这些概念有一定的了解。作为一个在某种意义上领先于时代的国家，新加坡可能是见证席卷全球的智能城市技术持续升温的代表。

方法

为了考察关于数据的使用体验和在其使用背后的原因，我们进行了最初的调查工作。从大量的新加坡人口中，政府雇员依据他们对科技的熟悉程度和他们在日常工作中对不同种类科技的接触程度被选中参与了研究。总共有3221位雇员被征询回答调查问卷，代表了约20％的新加坡政府员工。同时对决策制定时使用实时数据进行了定性分析，提供了一个基于情境的理解。项目的这个部分关注点是观察个人在横跨岛屿的几个不同枢纽车站里如何使用和搜集信息。这项研究提供了在研究中流动性部分的回答是如何转化为行动的洞见。

进行时

最有可能获取的可能是与流动有关的实时数据。在这点上，大多数受访者称“使用实时数据来为他们的流动和交通行程安排提供信息。”；超过三分之二的人会搜寻实时交通讯息。人们现在会为每天的行程计划搜集信息——包括公共交通和道路交通情况。

个人搜集数据并改变出行方式的原因主要分成三大类：钱，时间和

难易度（表4.3a）。对于大多数人来说，改变行程计划的首要目的是节约时间。其他经常被提及的原因是使行程更容易管理，优化他们的任务和差旅以及让整个行程更“顺畅”。为节约开支改变行程排在榜单第5位，我们可以猜测这是由于乘坐新加坡交通系统相对便宜。受访者们写下的评论也支持了这些发现；大多数参与者声称，因为节约时间或更快到达目的地而使用实时数据来改变通勤。

表4.3(a) 比较实时数据和非实时数据的用户：你的行程平均愉悦度如何？

项目	愉悦值（由数据得来）			预期愉悦值（由数据得来）			平均值差异
	平均值	数据个数	均方差	平均值	数据个数	均方差	
步行	3	34	0.7	2.2	20	0.48	-0.8
公共巴士	2.4	68	0.78	2.8	23	0.43	0.4
公共轨交（MRT/LRT）	2.7	70	0.7	2.4	34	0.51	-0.3
出租车	2.8	33	0.77	2.5	14	0.51	-0.3
私家车	3.2	43	0.83	2.6	24	0.5	-0.6
注释 1=不高兴 2=有点不高兴 3=中性 4=有点高兴 5=高兴							

【表4.3a】

为衡量对实时交通信息的定性反应，每个人被要求对他们使用不同交通方式的实时信息前后通勤的愉悦度进行打分。如果某个人不使用实时数据，则被要求推测自己的情绪反应来估计对于实时数据的期望值大小。正如在先前的文献中看到的，虽然公共轨道系统（MRT，Mass Rapid Transition，大规模捷运系统）和LRT（Light Rail Transition，轻型轨交系统）的改进最微弱，但人们回答，使用数据使得他们的通勤更愉悦。这很可能是因为火车到站时间很规律，间隔时间通常少于5分钟。在步行交通的情况里，预期的改善比实质感知到的变化要强至少100％。这不平衡的预期可能导致对公开数据倡议的不信任，尽管它们正按照计划执行。一个显著的例外是公共巴士的信息使用；现在不使用数据的人期待的改进比使用者看到的改进更微弱。

虽然如同在上文提到的先前研究中所见，实时信息能够影响通勤的愉悦度，我们发现了一个有趣的现象，个体不会每次都因为从数据中得到的信息而改变他们的通勤路径。有人推测个体也许会更加享受他们的

等候过程。当问到实时信息的收集是否会导致“在你交通出行时候有目的地改变等候时间”，57.5％的人回答是“有时”，16.5％的人回答“从不”，而19.7％的人回答“经常”，只有6.3％的人他们等候的时间总是会改变。与此相似，个人也不常改变他们出行方式的选择，例如从MRT转到巴士或者出租车。大多数人说到他们有时会转换出行方式（65.4％），而21.3％的说到他们从不改变。剩下的11.8％回答他们经常改变方式，只有1.6％的人回答总是会改变方式。

然而，这些发现看起来与早先的发现，即人们希望缩短等候与通勤时间相矛盾。对于出行方式转换这一问题，一个可能的原因是依赖于巴士系统的人数比依赖其他出行方式的人数多得多，多条线路的巴士在岛中央区域沿相同路线行驶。比如在人们公园综合（People's Park Complex）巴士站，有8条沿着余东璇路（Eu Tong Sen Road）行驶的巴士线路停靠，这也意味着即使当前信息提示乘客巴士服务出现延迟，选择乘坐巴士也是方便的。

传达的讯息

在交通方面，许多文献强调了站台实时信息标牌在改变公共交通等候时间与知觉可靠性方面取得了一定的成功（Al-Deek等人，1988：Boyce，1988，Dziekan和Kottnehoff，2007，Dziekan和Vermeulen，2006，Schweiger，2003）。但考虑到安装与实施不同获取信息途径的花费，人们从哪里获得数据才能让数据尽可能的被大众使用成为了一个突出的问题。创造不同的获取数据方案的困难在于人们在决策过程中使用数据的方式多种多样。

文献中提到了在哪里看到信息对改变与能源消耗相关的行为的重要性；在改变我们与资源使用有关的下意识行为上必定经常能看到数据的身影。这意味着数据所处的位置，以及信息获取的时间相对我们做出决策的时间点都很重要。对于交通数据的挑战在于，大量的人都提到，在他们做完行程规划后才会用到数据；数据用来帮助决定出行方式的转换

或者减少等候时间。在定性分析中个体给出了同样的回答。数据在决策过程的偏后部分才派得上用处。



【图1】

无论是在调查还是在观察中，智能电话和平板电脑在这个社会中的重要性都不能被低估，个体与他们的设备间有着清晰的联系。在任何公交车站或是在MRT上，发现大量的个体在用他们的设备与朋友聊天，玩游戏，上网等等。伊藤等人在日本发现了相似的文化现象，在公共交通中数字设备对调节社会生活起到了越来越重要的作用。特别地，设备在这样的环境下是用来观看的；打电话的举动已经越来越少，并且常被认为是粗鲁的行为[25]。

现在，政府不同机构间通过API来共享他们的数据，使得公众可以接触到数据，并且开发出大规模的工具。一个例外是与公共交通有关的机构。在2011年5月，两大公共交通服务提供商之一SBS在他们的服务中关闭了对开发者开放的API[26]。这一行为使得几个独立应用和服务无法使用。SBS留下了自己的Iris应用以及功能齐全的实时巴士信息。作为规划行程状态的工具，它获得了巨大的使用量，但是许多人指责它不直观、缺乏关联性、难以使用。在意识到受访者们压倒性地偏好把智能电话作为他们获取数据的窗口，有必要对数据开放和和对于多用户群组的

数据获取进行更多的讨论：相比于现在获取数据方式，更多关注应放在用户偏好的数据获取方法上。

创造和分享的实时信息可能来自多个不同的源头，每个源头都有它自己的优点和缺点。毫无意外的，这个调查群体高度信任政府来源的信息。在移动领域，环境质量和安全保密等方面，回答者提到平均来说大多数时间他们信任实时信息（表4.6a）。对非公共来源的信息，诸如来自电视或者广播新闻，也有一定的信任感，但是程度相对较少。进一步来说，对于从人群中获得的信息，尽管很多情况下是非常准确的，但是信任度在三个来源中是最低的。

表 4.6(a) 你有多信任以下的信息源?				
	项目	平均数	数据个数	均方差
移动性	官方公共信息（例如政府）	3.19	124	0.46
	非公共信息（例如私人电视新闻，非政府应用）	2.88	124	0.45
	公众来源	2.56	124	0.59
环境质量	官方公共信息（例如政府）	3.22	120	0.48
	非公共信息（例如私人电视新闻，非政府应用）	2.96	120	0.44
	公众来源	2.49	120	0.57
安全与保密	官方公共信息（例如政府）	320	121	0.5
	非公共信息（例如私人电视新闻，非政府应用）	2.96	121	0.46
	公众来源	2.54	120	0.62
注释 *1=总是不信任 2=大多数时间不信任 3=大多数时间信任 4=总是信任*				

【表4.6a】

个体改变他的下意识行为必须建立在相信信息的基础上，屡次错误的数据会导致不满和怀疑。然而，这项调查的受众可能也被机构雇佣并且参与创建政府提供的数据集，或者可能这些数据与他们在职业身份下接触到的数据相一致，因此他们更加信任政府提供的数据显得毫不意外。

当寻求个体行为的变化时，对数据的信任度这一命题变得很重要；而建立信任是很难的。调查的受访者对政府提供的实时数据信任感压倒性地超过其他的信息源。这可能是由于这些政府雇员对信息来源的熟悉程度，或者数据本身出自于其日常工作。然而，在定性研究中被询问的受访者多次表示出对实时交通信息的不信任。一些人提到了它只是政府

的一种宣传，而其他人则不认为显示的时间表与他们对巴士运行的直觉经验是一致的。

在计划与决策过程中采纳实时信息面临的一个主要挑战正是在于数据的实时特性。等候时间可能更久，或者由于未知的交通情况导致延长几分钟，这样的现实会让一般用户感到困惑和沮丧以及出现不可靠和不信任感。从某种程度上说，它类似于人们对天气预报员的普遍不信任：我们收听天气预报，却知道它很有可能是不准确的。可笑的是在定性研究中，许多人在使用站台上的实时信息标牌时参照了打印出来的汽车时刻表－衡量静态时刻表相对现实中实时信息的偏差。这一点在人民公园（People's Park）站年长的乘车者中特别显著。



图2

对实时信息不信任的情绪阻碍了对于支持获取实时信息的指示牌和基础设施进行的大额投资。令人惊讶的是，在站台的标牌上描述实时信息的标示，即使加上打印出来的地图，对解释那些实时信息也少有帮助。当这些信息在全岛范围内变得更容易获取的时候，也应该考虑教会更多的公众理解这些信息。



【图3】

体验数据是信赖数据的关键。有人也许会质疑，与“精通技术”有关人口的统计特征，例如教育程度和年龄，可能是信赖度的自变量。然而，在信赖度和教育程度或年龄之间并没有找到显著的联系。

结论

感知方面的测量挑战在于它们会随时间而改变。技术上的改变、对于特定设备愈加熟悉、日新月异的经个人承担的经济和社会压力都能在未来改变类似研究试验的发现成果。这项研究在当前的社会处境下、针对当前的技术和现在的市民，提供了一条衡量标准与一种对数据不同的感知。

具体地说，就数据而言“如果你实现了（技术），自然就有用户光顾”这一想法在新加坡执行得很好——很大程度上是因为新加坡拥有冒风险去创造跨模式获取信息工具的资本。然而，这项研究的洞见是，数据介入的地点可以更有效率和准确地瞄准特定的习惯与行为方式，聚焦于用户实质上使用岛上可以获取实时信息的广度。总体的趋势是通往开放数据，但针对改变用户行为这个期望来说，特别关注更有目标性和聚焦性的干预可能会比当前的“鸟枪”战术就数据获取而言更有效果。

参考文献

1．Al-Deek, H., Martello, M., May, A., Sanders, W. Potential Benefits of

in Vehicle Information Systems in a Real Life Freeway Corridor Under Recurring and Incident Induced Congestion. University of California, Berkeley: Institute of Transportation Studies, 1988.

2．D, Boyce. „Route guidance systems for improving urban travel demand and location choice." Transportation Research 22, Nr. 4 (1988):275–281.

3．Darby, Sarah. The Effectiveness Of Feedback On Energy Consumption: A Review For Defra Of The Literature On Metering, Billing And Direct Displays. Oxford, UK: Environmental Change Institute, University of Oxford, 2006.

4．Department of Statistics Singapore. Singapore Census of Population 2010: Statistical Release 2— Households and Housing. Singapore, 2011.

5．Dziekan, K., Kottenhoff, K. „Dynamic at-stop real-time information displays for public transport: effects on customers." Transportation Research Part A 41 (2007): 489–501.

6．Dziekan, K., Vermeulen, A. „Psychological effects of and design preferences for real-time information displays." Journal of Public Transportation 9 (2006): 71-89.

7．Golder, Scott, Michael Macy. „Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures." Science 333, Nr. 6051 (2011): 1878-1881.

8．iMerlion „The full ST interview with SG Buses developer Muh Hon Cheng".

iMerlion. http://www.imerlion.com/2011/05/full-st-interview-with-sg-buses.html (Letzter Zugriff am 6. August 2012)

9．Infocomm Development Agency. Technology and You: Singapore Infocomm Foresight 2015. Singapore: Government of Singapore, 2005.

10. Ito, Mizuko, Daisuke Okabe, Misa Matsuda. Personal, portable, pedestrian: mobile phones in Japanese life. Cambridge, Mass.: MIT Press, 2005.

11. Mahizhnan, Arun. 1999. „Smart cities: the Singapore case." Cities 16, Nr. 1: 13-18.

12. NCB. A Vision of an Intelligent Island: IT2000 Report. National Computer Board, Singapore. 1992.

13. Pike Research. 2010. „Smart Cities". http://www.pikeresearch.com/research/smart-cities (Letzter Zugriff am 4. August 2012).

14. Transit Watch - bus station video monitors customer satisfaction evaluation : metropolitan model deployment initiative: Seattle, Washington. Seattle, WA: Battelle Memorial Institute, 2000.

15. Williams, A., Robles, E., Dourish, P. 2008. „Urbane-ing the City: Examining and Refining the Assumptions Behind Urban Informatics". In Foth (ed.), Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City. Hershey, PA: Information Science Reference, IGI Global.

追踪垃圾，减少浪费

作者　大卫·李（David Lee）

译者　陈然

导论

城市感知技术最有前景的应用之一是向城市居民展示数据，从而揭示城市中平常所看不见的变化过程。这可以是诸如空气质量变化或附近道路的交通拥堵情况这样的短期过程，也可以是像增长的犯罪率和气候变化这样的长期过程。每一种应用都是加强城市居民对某个问题的意识并且唤起他们某种反应的机会，不论是通过转变他们的观念，还是改变他们的行为，这可能会激起更多讨论和积极行动。

对大多数城市而言，新建一套电子传感器基础设施的成本过高。一种较节约成本的替代方案是参与式感知，即由许多志愿者独立完成感知任务，并借助他们的数据勾勒出完整图景（Paulos、Honicky和Hooker，2008）。例如，市民在街上发现枯萎的树木，拍下照片，然后发布在自己的博客上，其成本要低于在每棵树上安装传感器来实时监控它们。

这些方法让更多的人参与其中，使得感知过程更为透明，且威胁性也可能更低（Burke等人，2006）。此外，尽管对科研或政策研究而言，由非专业人士采集的数据可能不够权威，但它更适用于对市民的反馈（Dutta等人，2009）。例如，如果想要报告空气污染对行人的影响，那么我们应在行人最密集的街道上检测，而不是在路灯或大楼顶上。此外，我们对另一个问题的理解也还不够充分，即如果我们参与了数据收集过程，那么我们看待数据的视角将会有多大差异。让市民帮助测量那些几乎看不见的现象（如低层大气中上不断恶化的臭氧污染），所得到的结果，相比仅凭专家所公布的，更令他们信服。而且，投入了时间去揭示问题的市民更可能改变他们的行为。然而，参与过程中所暴露出来的第一手资料的限制或缺陷，可能让我们对得到的数据更缺乏信

心。这些都会对城市感知产生深远影响。

垃圾追踪

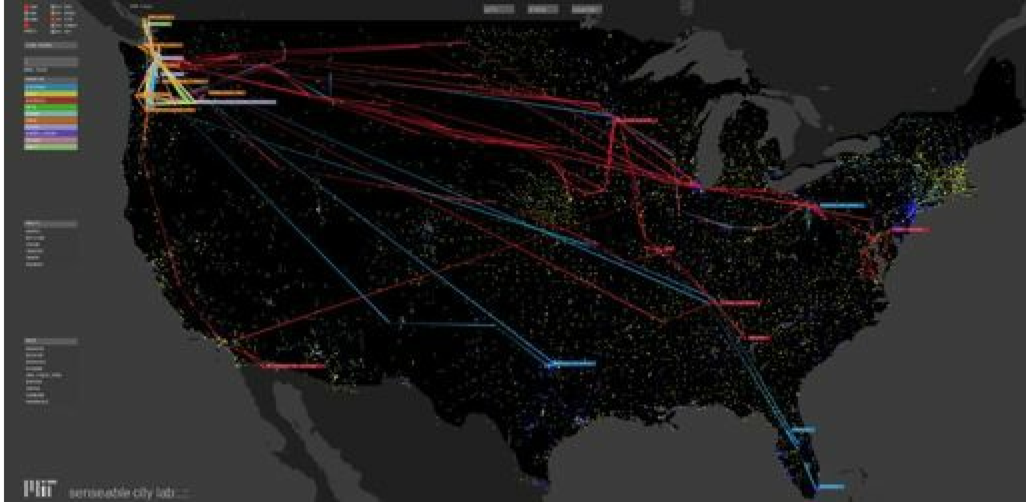
垃圾追踪计划源于一个简单的想法——其目的是了解垃圾在离开我们的视线后究竟去了哪里。现在的技术使我们能够追踪任何人造物从最初创造到最终处理过程中的移动轨迹，管理垃圾处理结果的系统极其复杂，很少有人了解。垃圾的运输方式有卡车、火车、轮船和飞机等，其移动路线是一个由垃圾中转站、垃圾填埋场和再处理厂构成的设计精巧的网络，并应法规和市场要求不断变更路线。我们希望了解这些基础设施的完善程度，也想知道，像丢弃一部手机这样的个人行为如何造成大范围影响。

2009年10月，麻省理工学院可感城市实验室使用远程跟踪传感器，通过西雅图垃圾处理系统，对垃圾运动轨迹进行了研究。实验室的研究人员前往西雅图，部署了2000个GPS追踪“标签”；每样垃圾都被贴上标签，然后被散布在城市各处（Boustani等人，2011）。我们通过垃圾追踪网站在西雅图市民中招募志愿者；入选的志愿者需要提供垃圾，帮助给标签做目录，将标签贴到垃圾上，然后使垃圾进入旅行过程。图1照片为我们与当地志愿者一起做准备工作时的场景。

【图1】志愿者在帮助做分布式传感：识别可追踪的垃圾，使用具防护性的速凝泡沫胶给垃圾贴标签，然后将其丢弃。

在接下去2个月内，我们对每一件垃圾的位置进行远程跟踪，通过轨迹图、动画和视频的形式使之可视化，并在公开展览中和网站上展示结果。这使得志愿者和普通大众都能实时看见垃圾被丢弃后的“第二生命”。从这些轨迹图中我们清楚地看到，垃圾的旅程远远超出城市范围，它们跨越州界，甚至国界，尤其是电子垃圾和危险废弃物。垃圾的运输方式和去向可能对其所带来的能源消耗和排放的影响有着巨大差异。

我们试图通过与垃圾“对话”的方式，将市民与无处不在却又难以捉摸的城市基础设施联系起来。展示看不见的垃圾处理过程能增强市民对个人所排放的垃圾对环境影响的意识，并鼓励他们采取更可持续化的生活方式。这些信息会影响城市、地区或全球范围内的基础设施规划，以及个人的垃圾处理决策。


【图2】实验中每件被追踪垃圾的可视化位置信息。

在研究中，我们调查了参与这个实验的志愿者，以估量他们对实时垃圾运行轨迹图的理解和反应方式。我们也调查了非志愿者，以确定他们的反应是否有所不同。我们希望知道：帮助部署传感器的经历是否改变了志愿者对这些结果的看法？如果我们希望大规模地利用垃圾追踪计划来改变人们的态度或行为，让他们参与其中是否会有帮助？

方法

在2009年10月的两个周末，共有84名志愿者参与了垃圾追踪项目。在标签部署任务完成后，我们仍与每位志愿者保持联系。他们组成了研究对象A组。

B组由访问垃圾追踪网站的普通用户组成。由于他们在查看这些数据之前没有参与过任何垃圾追踪项目，所以我们将他们作为对照组。他们来自与最初招募志愿者相同的社交网络，通过电子邮件发出的网站公告招募而来。

追踪结果的网站界面

我们以带时间戳的地理坐标的形式，收集了标签垃圾三个月内的实时轨迹。我们用原始轨迹在谷歌地图中叠加图层，用自定义坐标表示垃圾的第一个和最后一个位置，用线条勾勒出这两个位置中间所有的点。用户可以单独查看每一样垃圾的轨迹，也可以批量查看垃圾的最新位

置。用户还可以根据垃圾类型进行筛选，或只查看他们自己的垃圾（针对提供垃圾的志愿者）。图3是两个基本的网站界面视图。由于显示的数据是实时生成，在紧接着初步部署的几个月内，实际显示结果会有变动。



【图3】向公众展示垃圾追踪数据结果的网站界面示意图。左图是单件垃圾的轨迹图；右图是包含许多标签最终位置的聚合视图。白色标记代表垃圾运转站和西雅图的垃圾处理场。

我们要求两组参与者均完成两项调查，一项在查看追踪结果网站前，另一项在查看后。调查问题试图捕捉个人对垃圾追踪数据的反应方式。两项调查都采用李克特量表的形式，问题涉及调查对象当前关于其决策对环境所造成的影响的态度和信念，包括对垃圾产生和处理的态度，以及采取可持续性的行为的频率，例如垃圾回收和把电子垃圾送至指定的处理场。

我们检验了如下两个假设：

1．不论是否查看过垃圾追踪数据轨迹，志愿者和非志愿者对垃圾处理相关问题均持不同态度，表现出不同行为。

2．查看垃圾追踪数据会引起态度和行为的改变。但对志愿者和非志愿者而言，改变的实质有所不同。

结果

在网站用户能够查看标签垃圾的轨迹数据之前，我们使用在线调查工具调查了200多名访问者。一个月后，我们联系了每一位参与者，邀

请他们参与后续调查，其中70名初始参与者完成了调查问卷。在这70份回应中，有32份来自西雅图部署计划的志愿者，其余38份来自网站访问者，他们并不在志愿者之列。

我们首先比较了两组群体的一般反应。志愿者对当地垃圾处理和回收系统信心更足，也清楚了解垃圾被处理后的去处。在查看垃圾追踪数据后，他们也试图减少垃圾排放，参与更多垃圾回收，且参与程度显著高于非志愿者。相比之下，查看追踪数据缩小了两组群体之间的差距。尽管在最开始时，志愿者群体似乎对了解危险废弃物处置场的位置和垃圾追踪的工作原理更有信心。

有序逻辑回归模型进一步揭示了志愿服务与对可持续性问题更强烈的反应的关联程度。在购物时，志愿者更有可能考虑产品包装所产生的垃圾。他们更倾向于相信垃圾处理系统的效率和路边回收的成效。相同的模型显示，在一般情形下，查看数据轨迹只改变了调查对象对调查中两个问题的反应。在控制了垃圾追踪轨迹的不同方面后，两组群体对垃圾去处及垃圾追踪系统的工作原理都有了更深了解。

最后一个回归模型检验了在这两次调查期间，志愿者改变反应的方式是否与非志愿者有显著不同。该模型揭示了以下三点：

在事前调查中，志愿者比非志愿者更有可能了解危险废弃物处置场的位置。

在查看垃圾追踪轨迹后，非志愿者更有可能知道危险废弃物处置场的位置。

志愿者在查看垃圾追踪轨迹后的反应相反。在事后调查中，他们反而不太知道危险废弃物处置场的位置。

结论

总而言之，这些模型显示，尽管从垃圾追踪中得到的信息反馈的确显著改变了部分人对垃圾问题和感知技术的理解，但它没有显著改变他们对垃圾产生和垃圾处理的态度和行为。

参与实验的志愿者更频繁地参与有利于环境可持续发展的行动。平均而言，当不在家时，他们比起非志愿者更多地尝试以重复使用、减少包装和垃圾回收等行为来减轻对环境的影响。他们对垃圾处理系统的有效性和自己对它的了解程度都更有信心。然而，这些影响在长期中并非一直有效。志愿者享受完参与实验的“快感”并看到成果后，在随后几个月内，他们从中得到的影响逐渐消失殆尽。除了首次参与外，再也没有新的信息获取或其他参与途径，志愿者对该事件的记忆可能随着时间退却。两次调查之间的三个显著变化都出现在以知识为基础的问题上。在关于对垃圾去处和垃圾追踪项目的工作原理的了解程度，以及危险废弃物处置场位置的问题上，志愿者和非志愿者都呈现出明显变化。作为意在唤醒环保意识的一项实验，垃圾追踪项目在解释垃圾处理过程及追踪技术如何用于揭示其路径上取得了一些成功。

与此同时，志愿者和非志愿者对后一个问题（了解危险废弃物处置场位置）的反应都发生了变化，但变化方向恰恰相反，这表明了查看追踪轨迹的另一点作用，也是具有潜在积极效果的作用：对某些人而言，这些轨迹解释了危险废弃物处置场所的问题，而这正是他们此前所不知道的；对其他人而言，追踪轨迹表明，该系统的复杂程度和不可预知程度远远超出他们此前的想象，这迫使他们重新评估对自己所生活的这个城市的确切了解程度。

考虑到自我选择的内在问题及应用自愿调查的天花板效应，在从这个研究中得出普遍结论时，我们必须非常小心。在垃圾处理方面，我们必须使用更大的样本数和更多的控制设置，更深入地研究这些效应，以了解参与式感知如何影响全市范围内的行为改变。从调查的开放式问题中，我们发现了一些有趣的事，代表了一些个体的态度和行为转变：

垃圾追踪使我更了解某种特定物品的处理细节。此前，我懒得去查该把旧电脑或灯泡送去哪里处理，我也不想把它们随便扔掉，于是它们就一直在我身边搁置着。垃圾追踪项目让我有动力去了解这类物品的正

确处理过程。

在我的垃圾及其去处之间画出的轨迹线令我产生一种责任感。希望这能帮助我缩小消费选择与保护地球环境之间的差距。

我们相信，这些引用的例子反映了参与式感知项目的潜在影响力，未来的项目应尽可能稳定地产生持久的变化。例如，关注电子垃圾和它们有时较长期的、不可预知的轨迹图能使参与者参与更长期的项目。不安全的电子垃圾处理方式产生的风险更迫切也更严重，而市民也往往有更多可供追踪与比较的处理方式选择。一个在全球范围内实施的、由参与者在家独立部署的众包电子垃圾追踪项目，可能会为集体产生有价值的数据，同时向个人展示垃圾处理的最佳方式。

幸运的是，感知垃圾位置的技术十分先进，且越来越具适用性，能够为市民产生更精确的实时反馈。但这些反馈如何能鼓励可持续性的行为并转变成社会规范，则值得进一步仔细研究。

致谢

我们要感谢在西雅图帮助我们执行初步实验的垃圾追踪团队和志愿者；高通公司、废物管理公司、斯普林特公司、西雅图市（西雅图艺术和文化事务办公室、西雅图公共事业部门、西雅图公共图书馆），以及感知城市实验室的所有成员，感谢他们在物质和技术上提供支持。

参考文献

1．Boustani, A., L. Girod, D. Offenhuber, R. Britter, MI Wolf, D. Lee, S. Miles, A. Biderman, and C. Ratti. 2011. "Investigation of the Waste-removal Chain Through Pervasive Computing." IBM Journal of Research and Development 55 (1.2): 11-1.

2．Burke, Jeffrey A., D. Estrin, Mark Hansen, Andrew Parker, Nithya Ramanathan, Sasank Reddy, and Mani B. Srivastava. 2006. "Participatory Sensing" (May 5).

3．Dutta, Prabal, Paul M. Aoki, Neil Kumar, Alan Mainwaring, Chris

Myers, Wesley Willett, and Allison Woodruff. 2009. "Common Sense: Participatory Urban Sensing Using a Network of Handheld Air Quality Monitors." In Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems, 349-350. SenSys' 09. New York, NY, USA: ACM.

4．Paulos, E., R. J. Honicky, and B. Hooker. 2008. "Citizen Science: Enabling Participatory Urbanism." Handbook of Research on Urban Informatics: 414–436.

纽约对话交流：通过全球电信展示城市动态

作者　弗朗西斯卡M. 罗哈斯（Francisca M. Rojas）

译者　姚琪

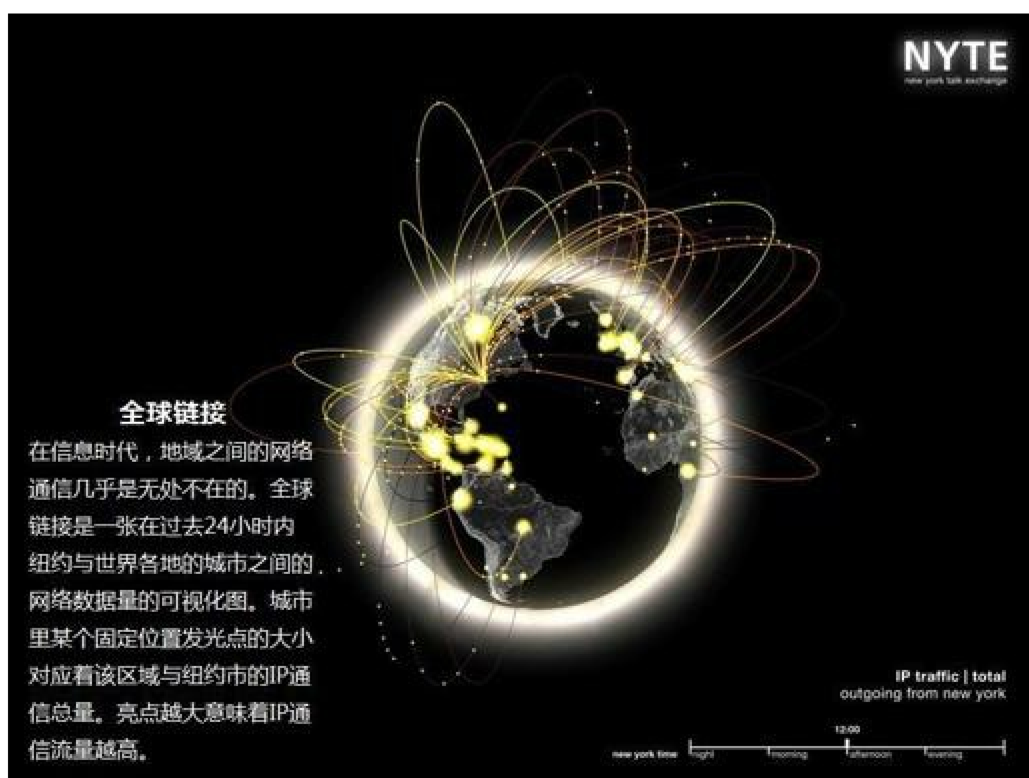
导论

这项研究是基于麻省理工学院感知城市实验室（Senseable City Lab）的纽约对话交流项目。该项目根据2008年9月纽约市与世界其他地方的电话数据生成了多种可视化效果（罗哈斯等人，2008）。本研究进一步分析了这些数据，以便更好地理解电信流是如何揭示全球城市原本无形的动态趋势。通过电信网络，纽约跟世界上的哪个地方联系最紧密？通过对国际电话的估算，可以认为哪些街道最具全球性？以及这些全球电信链的目的何在？

用电信数据来理解城市动态，并以此来指导城市规划和政策，这种方法并不新鲜。伊锡尔·德·索拉·普尔（Ithiel de Sola Pool）指出早在1915年，“电话公司就是研究城市系统化人口统计信息、街道发展趋势和特点的主要来源。”（1983: 42）近一个世纪以前，这些数据的使用引发了城市规划领域的扩张。20世纪50年代中期，地理学家戈特曼（Gottman）也认识到从信息流中可以挖掘出有价值的东西，来进行城市研究。戈特曼分析了AT&T在大都市纽约（1961年）的国内长途电话数据，根据各地区之间的信息链的密度和走向，我们可以掌握来自这些地方的活动。“电话流密度可以很好地衡量人际关系，而人际关系往往跟地区经济挂钩，所以电话不仅体现了经济关系和政府关系，还体现了社会和家庭的联系。”（戈特曼：590）。半个世纪之后，由城市形成的数字痕迹（digital trace）使我们能够采用这种方法，在全球范围内展开对这种联系的调查。


NYTE
new york talk exchange
世界在纽约
世界在纽约展示了通过AT&T电话网络，不同的街道与世界其他地方的联系有着很大的差异。将城市划分成了一系列由像素块组成的网格，根据世界各地区中与纽约联系最紧密的城市，给每个像素块分配了的不同颜色。色条的高度代表世界各地区与每个街道通信所占的比例。每个像素块都代表着一个城市，这个像素表对与纽约交流最多的城市进行了排行。
The Bronx
Manhattan
Queens
Brooklyn
Staten Island
City Ranking
Flushing, Queens

【图1】


NYTE
new york talk exchange
全球链接
在信息时代，地域之间的网络通信几乎是无处不在的。全球链接是一张在过去24小时内纽约与世界各地的城市之间的网络数据量的可视化图。城市里某个固定位置发光点的大小对应着该区域与纽约市的IP通信总量。亮点越大意味着IP通信流量越高。
IP traffic | total
outgoing from new york
12:00
new york time
night
morning
afternoon
evening

【图2】

数据

本研究中的分析依据三种信息来源：1）由美国一家主要的电信运营商提供的长途电话记录集，2）来自美国人口普查局的街道人口统计数据，3）实地调研包括在选定的街道中开展的半结构化（semi-structured）采访和观察。呼叫数据包括了纽约市各街道与全世界超过200多个国家的长途电话话务量。两地之间的通话次数和总的通话时间以每10分钟一个梯度的方式集中起来，统计了2008年9月每天中产生的数据。这包括了固定电话、无线电话和预付费电话卡等方式的电话。在纽约，电话是按照地理位置分布的，以一个线路中心或电信交换站的规模聚合。为了分析研究，我们将城市划分成了67个不同的区域。在相同的地理范围内，电话数据会与人口数据相结合。我们估计，在这项研究中所分析的数据约占了纽约国际话务总量的25％[27]。这就是在本研究的分析中用作归一化呼叫量的数据。

上曼哈顿和中央皇后区有高密度的国际电话，我们在这两个地区进行了24次半结构化的采访。采访的问题跟呼叫数据有直接的关系。被采访对象要回答一个月通话次数、通话的时间长度、通话方式以及通话成本等问题。这种方法便于解释利用定量分析所得到的结果，可以让一些问题更加明确，诸如，用分钟来计算人均通话时间是否合理，每周一次的电话模式是否符合人们的生活经验。

一种双重地理的对话

之前分析的全球城市国际惯例表明，全球化中与纽约市的对手应该是它的主要商业合作伙伴：伦敦（英国），东京（日本），法兰克福（德国）（Sassen，2001、2002；Derudder，2008）。一些关于全球城市的研究中讨论了信息通信技术（information and communications technology：ICT）特别是廉价的电话，如何成为了连接移民和他们原籍国的“社会黏合剂”，来帮助移民进行跨国活动（Levitt，2004；Vertove，2009）。通过分析从世界上最具有商业性和移民性的城市拨

出电话的被叫地点，我们期望纽约的电信流同时连接着其全球商业伙伴和其相应的移民输出国。

我们调查了纽约长途电话，并分析了它们的国际目的地，发现在纽约排名前10的呼叫对象中，7个是移民输出国，而只有3个是很重要的商业伙伴。图1表示按照国家划分，纽约的国外出生的人口和美国贸易伙伴的呼叫量比例[28]。多米尼加共和国和墨西哥分别排第一和第二，而且在纽约国外出生的人口中，这两个国家也是移民输出最多的。与此同时，以美国贸易数据为例，英国、加拿大和德国与纽约有着很明显的商业关系。出现在这个列表上的第三类国家跟美国有着贸易关系，而且会向纽约输送移民：墨西哥和印度。从这些国家的人均GDP来看，除了墨西哥和印度，强大的经济体倾向于跟美国进行商品贸易，而疲软的经济体倾向于向美国出口劳动力。

排名	国家	世界地区	呼叫地来自纽约 (%)	呼叫地来自美国 (%)	纽约外国出生人口 (%)	美国进口合作伙伴 (%)	美国出口合作伙伴 (%)	人均 GDP (美元)	联系的类型
1	多米尼加共和国	中美洲/ 加勒比地区	11.7	2.3	13.1	0.2	0.5	4,179	M
2	墨西哥	北美	9.1	17.8	6.4	10.8	12	9,516	M/T
3	英国	欧洲	7.5	4	1.2	2.9	4.4	45,510	T
4	加拿大	南美	7	12.2	0.8	16.1	12.9	43,396	T
5	危地马拉	中美洲/ 加勒比地区	5.5	3.2	0.8	0.2	0.4	2,548	M
6	厄瓜多尔	南美洲	5.3	0.8	4.9	0.3	0.3	3,432	M
7	牙买加	中美洲/ 加勒比地区	3.3	1.1	6.3	0	0.2	4,565	M
8	印度	亚洲	3	10.8	2.7	1.3	1.6	947	M/T
9	德国	欧洲	2.5	2	0.8	4.9	4.4	40,308	T
10	菲律宾	亚洲	2.4	3.5	2	0.5	0.7	1,624	M

根据2008年9月份纽约市总的国际通话时长，列出排名前10的呼叫地，并与美国2007年总的国际通话时长、来自原籍国的纽约外国出生人口（2006-2008）、美国交易伙伴（2007年）和人均GDP（2007年）做了比较。

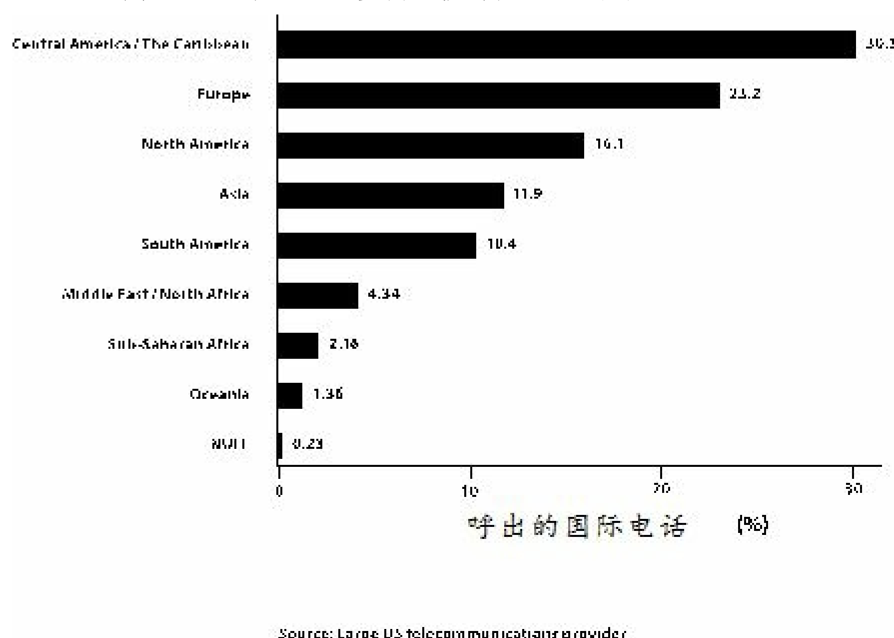
注意：M= 移民, T= 贸易, M/T= 移民和贸易伙伴

来源：作者采编的数据来自美国一家大型的电信运营商、美国人口统计局2006-2008年美国人口社区普查（ACS），2009年美国联邦通信委员会、国际货币基金组织2008年贸易统计指南、国际电信联盟2007年基本指标

【表1】

尽管根据国家调查的电话呼叫量会得到很丰富的信息，但可能会掩盖纽约对于世界上某一特定地区的吸引力。对世界各地区的呼叫量的比较（图2）表明，纽约跟中美洲和加勒比地区有着很强的联系，这反映了该城移民的人力[29]。从基于电信的全球链接看，纽约的趋势不同与美国全国，主要面向中美洲和加勒比地区（占国际电话总量的30％），其次是欧洲（23％）和北美（16％）。而美国作为整体而言，

国际电话往来通常是北美（29％）、亚洲（20％）和欧洲（15％）（FCC 2008），符合国家主要贸易伙伴的比例（IMF 2008）。



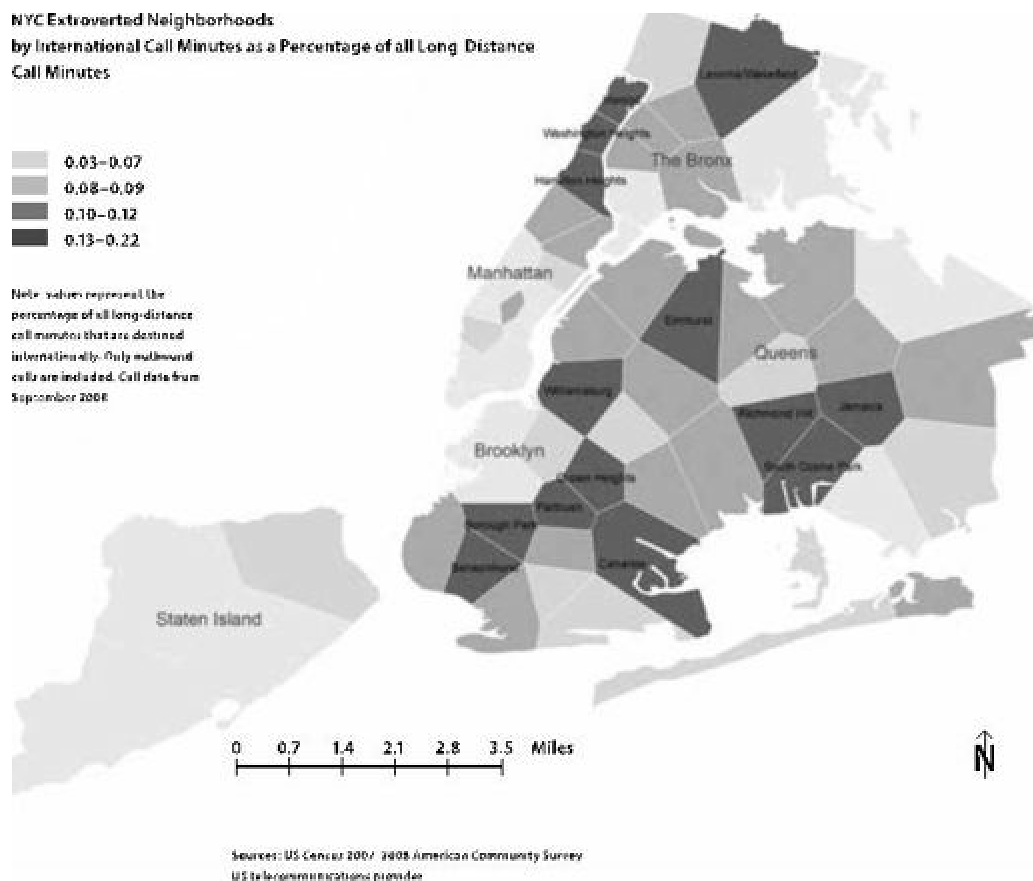
【图3】

纽约的呼叫地表现了商业伙伴和移民输出国的一种融合，因此揭示了国际电信联系的一种双重地理。根据电信可作为一种跨国活动工具的设想，对于纽约拨出的大量国际电话来说，移民所做的贡献和跨国企业不相伯仲。

移民者和国际电话

通过查看纽约市5个区的特定国外出生群体的集中程度是否与拨向不同国家的电话量相匹配，可以进一步验证移民者对城市国际电话流量的影响［30］。这种分析根据区的划分（布朗克斯，布鲁克林，曼哈顿，皇后区，史坦顿岛），对由每个区拨向目的地国的电话与国外出生人口的原籍国之间的关系做了简单的回归分析。为了能够让这两个变量符合正态分布并且在它们之间建立起线性关系，这个回归模型对两个变量进行了对数变换。

曼哈顿的回归模型拟合性最强，该区国外出生人口模型占了67％的



【图4】

国际通话目的地通话时长（2008年9月）与纽约市每个行政区中外国出生人口的OLS回归系数

行政区	布朗克斯	布鲁克林	曼哈顿	皇后区	史坦顿岛
回归系数	0.831*	0.964*	1.026*	0.984*	0.840*
	-0.093	-0.099	-0.09	-0.107	-0.157
常数	1.106	0.134	2.675	0.23	0.802
R平方值	0.55	0.59	0.67	0.56	0.32
N	67	68	69	69	62

注意：括号中的数值是标准误差。

*p < 0.001。

来源：作者分析的数据来自美国一家大型的电信运营商和美国人口统计局2006-2008年美国人口社区普查。

【表2】

曼哈顿的回归模型拟合性最强，该区国外出生人口模型占了67％的

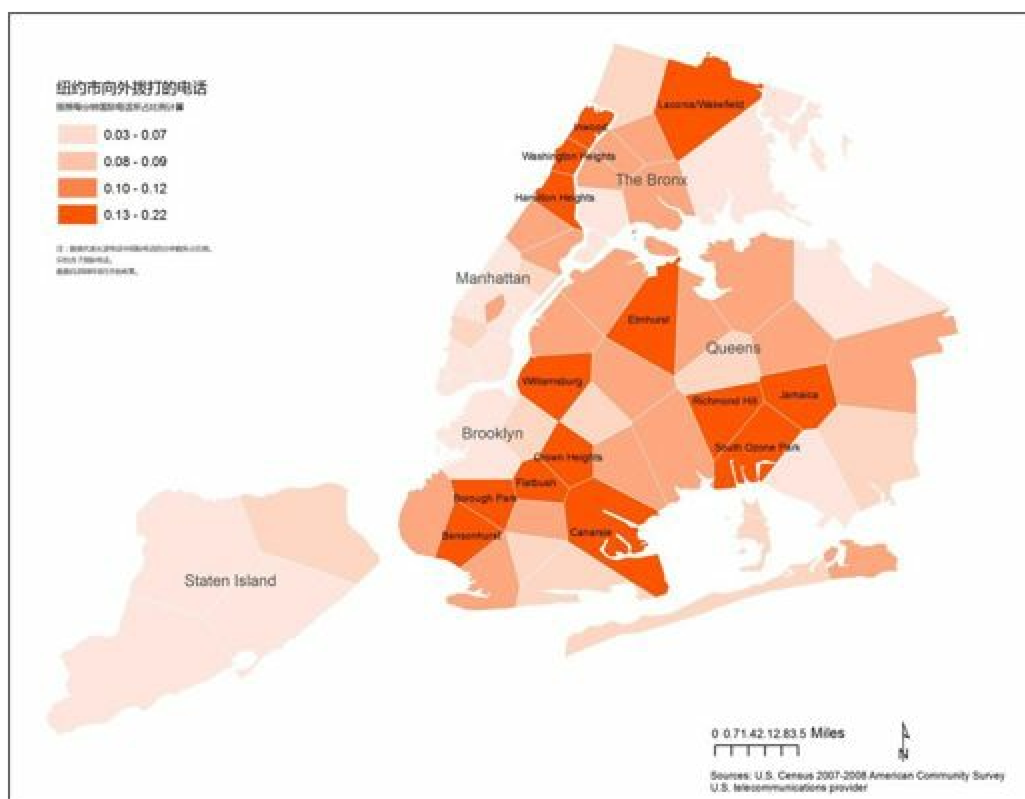
国际呼叫量。我们发现如果来自某个国家的曼哈顿国外出生人口增加10％，每个月从曼哈顿拨到这个国家的电话通话时间预期平均会增加10％。这种预测也同样适用于布鲁克林和皇后区。虽然这些结果几乎暗示着在区一级别上，国外出生群体与拨向他们的原籍国家的电话呼叫量存在一对一的关联，但是，由于忽略了原籍国家的收入水平，这些结果没有预想的准确。然而，国外出生群体和呼叫地国家在统计上都能够支持这样的假设：通信流量会表现出一个子城规模下的城市动态。这表明城市规划中建立移民原籍国和他们的呼叫地之间的关系，存在着潜在的价值。

在街道范围内，用电话数据来预示城市人口变化和经济增长，这种城市规划实践已有百年之久。找到的关于电信流的来源和目的地之间的关系表明，这种规划实践还可能得到进一步发展。在这个城市的功能和管理越来越趋向数字化的时代——带有GPS功能的交通系统已经可以实时掌握一辆公交车的具体位置——而且这个时代强大的计算能力可以很容易分析这些形成的数据集，城市功能和城市管理之间的信息循环瞬间就能实现。可想而知，纽约市城市规划局人口司可以根据国际电话的来源和目的地的数据生成估算算法，以定期预测在街道范围内的城市移民的流动状况。考虑到移民对于城市活力的重要性，这种方法可以加快城市普查工作而无须完全依赖美国人口普查的结果。很显然，人口普查不会计算国外出生的人口，尤其不包含低于区级别非法移民的信息。

外向的街道

为了确定城市的哪些区域是最面向国外的，或者说是“外向的”（Massey，1994），我们从区级别进一步降到了街道级别。外向的街道是那些国际通话时长相比于国内（美国）通话时长比例最高的街道。因此，这种分析涉及区分长途通话时间中的国内部分和国际部分，这两者组成了总的长途通话量。然后，我们比较纽约街道国际通话时长所占的比例与美国全国的国际通话占的比例，得出一个数值，以此来确

定是否将某一街道指定为外向的。根据2008年美国联邦通信委员会和国际电信联盟的数据，那年美国有12％的长途电话是国际的。根据这个阈值，我们对纽约64个线路中心中的14个进行了分析，国际电话呼出比例要高于美国平均值。



【图5】

外向的街道都是位于外区（outer borough）的城市主要住宅区，“……那里都是城市的工薪阶级和下层—中层阶级居住的地方，也是新移民定居的地方。”（Sanjek 2000: 29）确实，除了一个街道之外，其他所有街道在国外出生的人口比例都要高出城市平均的34％。皇后区的艾姆赫斯特地区，是其中一个更加外向的街道，在这些地方，全球性会变得更加明显。在惠特尼大街像到了印度尼西亚，而巴克斯特大街就像在哥伦比亚；印度，巴基斯坦和墨西哥的移民则集中在了罗斯福大街；沿着百老汇大街可以看到亚洲的缩影。

值得注意的是，没有一个城市的商业中心——不是指金融区或者市

中心——通过衡量国际与国内的呼叫比例就能被看做一个外向区域。然而，一些面向商业的线路中心，包括切尔西和金融区，是街道生成国际通话量最多的地方[31]，不管是就绝对数量还是就相对数量而言。

纽约的商业区				纽约的居民区			
街道	行政区	国际通话时间（分钟）	国内通话时间（分钟）	街道	行政区	国际通话时间（分钟）	国内通话时间（分钟）
切尔西区	曼哈顿	5.53	4.95	法拉盛区	皇后区	1.18	0.83
金融区	曼哈顿	2.24	2.67	华盛顿高地	曼哈顿	0.62	0.19
默里山区	曼哈顿	1.46	2.99	茵伍德区	曼哈顿	0.57	0.16
时代广场	曼哈顿	0.93	0.63	艾姆赫斯特区	皇后区	0.57	0.16
曼哈顿西区	曼哈顿	0.72	1.4	科罗娜区	皇后区	0.53	0.4
曼哈顿东区	曼哈顿	0.49	0.55	弗拉特布什区	布鲁克林	0.5	0.15
炮台公园区	曼哈顿	0.38	0.5	威廉斯堡区	布鲁克林	0.5	0.19
基普斯湾	曼哈顿	0.32	0.41	列治文山区	皇后区	0.49	0.16
鲍林格林区	曼哈顿	0.28	0.33	阿里斯托亚区	皇后区	0.47	0.28
克林顿区	曼哈顿	0.2	0.41	牙买加区	皇后区	0.47	0.18

用归一化的国际和国内呼叫的通话量排列了纽约市最外向的区域，并区分了线路中心和街道的类型：主要商业区或居民区。

注意：国际通话时长用人口数量做了归一化，估计了该美国电信运营商的市场份额，并根据所有线路中心的人均通话量做了索引。

来源：作者分析的数据来自美国一家大型的电信运营商。

【表3】

街道特征

街道环境对预测这个地方的国际关系力量会有多大程度的帮助呢？一些主要变量可以解释不同社区在电话上所表现出来的外向程度的差异：国外出生人口比例，会讲一门外语的人口比例，以及中等家庭收入。如果国际电话是移民跨国主义的“社会粘合剂”，那么高比例的国外出生人口和外语使用者可以看作预测街道外向电话活动的重要因子。我们确实发现在一个区域的国外出生人口比例是预测国际电话占总的长途电话时长比例的重要因子。

根据来自中等家庭收入的长途电话的比例，所展开的一种国际电话回归分析表明，收入跟国际电话的比例成反比关系。图6所示的回归系

数，采用了虚拟变量来对照收入分类——低等（38,320美元以下），中等（间于38,320－71,850美元），高等（71,850美元以上）——此时保持国外出生比例、外语使用者比例和每个街道的家庭数量不变[32]。忽略的变量代表了纽约中等收入街道。

变量解释	模型 1	模型 2	模型 3	模型 4
低等家庭收入	0.04**	0.04**	0.04**	0.04***
(低于 38,320美元)	-0.02	-0.02	-0.01	-0.01
高等家庭收入	-0.04*	0	0	0
(高于 71,850美元)	-0.02	-0.02	-0.02	-0.02
外国出生人口比例 (%)		0.20***	0.23***	0.22***
		-0.04	-0.06	-0.06
外语使用者比例(%)			-0.04	-0.04
			-0.05	-0.05
家庭数量				2.59^{-07}
				(2.11^{-07})
常数	0.09**	0.02	0.03	0.02
	-0.01	-0.02	-0.02	-0.02
R平方值	0.22	0.44	0.44	0.46
N	60	60	60	60

国际通话时间占总的长途电话比例与纽约市街道的低等和高等家庭收入、外国出生人口比例、外语使用者比例和家庭数量的OLS回归系数

注意：括号中的数值是标准误差。

*p<0.05，**p<0.01，***p<0.001。

来源：作者分析的数据来自美国一家大型的电信运营商（2008年9月）和2007-2008年美国人口社区普查（ACS）

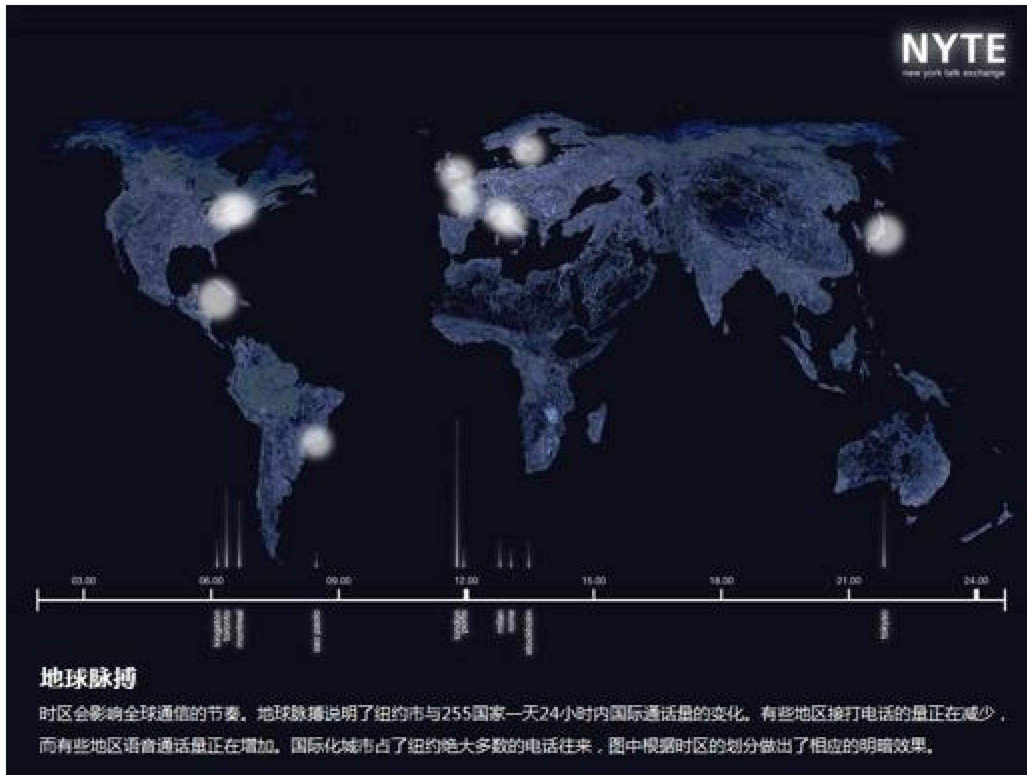
【表4】

所有模型表明，低收入街道的国际电话量预计会比中等收入街道高4%。在这方面低收入街道和中等收入街道有着显著的差别。但是高等收入街道的国际电话量预计会低于中等收入街道的4%。该结果只在模型1中有意义，但该模型没有控制包含在回归分析中的变量。该回归分析表明，国际电话时长的比例在不同方面受到国外出生人口比例和该区域中等家庭收入的影响。考虑到ICT在移民生活中的重要性，预计会发

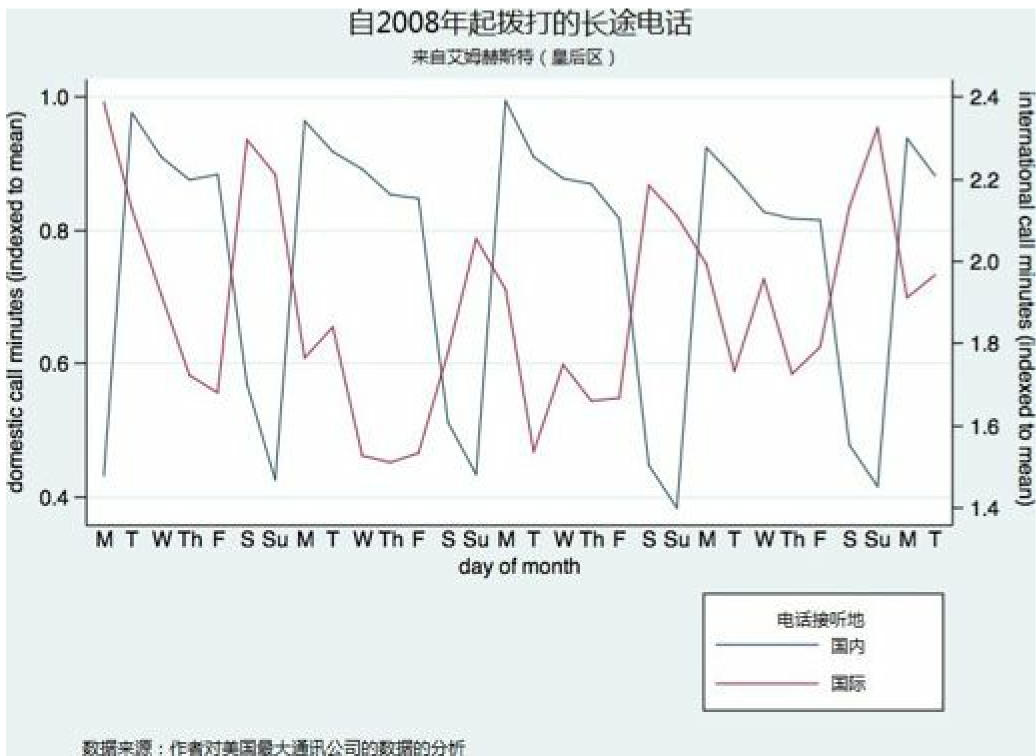
现一种国外出生人口和国际电话之间的正比例关系——这一发现进一步得到这项研究中的实地采访的支持——然而令人惊讶的是，收入和国际电话之间存在着明显的反比例关系。这意味着居住在经济资源最少的街道的人正是那些利用电话进行大多数跨国活动的人。虽然这部分要归因于国际电话费用的大幅下降，但这也证明了对于那些外向街道的移民社区来说，全球通信联系是何等重要，但外向街道上几乎所有的中等家庭收入都低于城市的平均值479,00美元。事实上，我们甚至可以预计中等收入区域所进行的国际通话要高于城市最高收入街道，包括商业区。然而，如图5所示，商业区占据了绝大多数的全球通话，但考虑到国际通话占总通话的比例，我们发现城市的低收入、移民区域的最经常进行国际通话。

通话模式

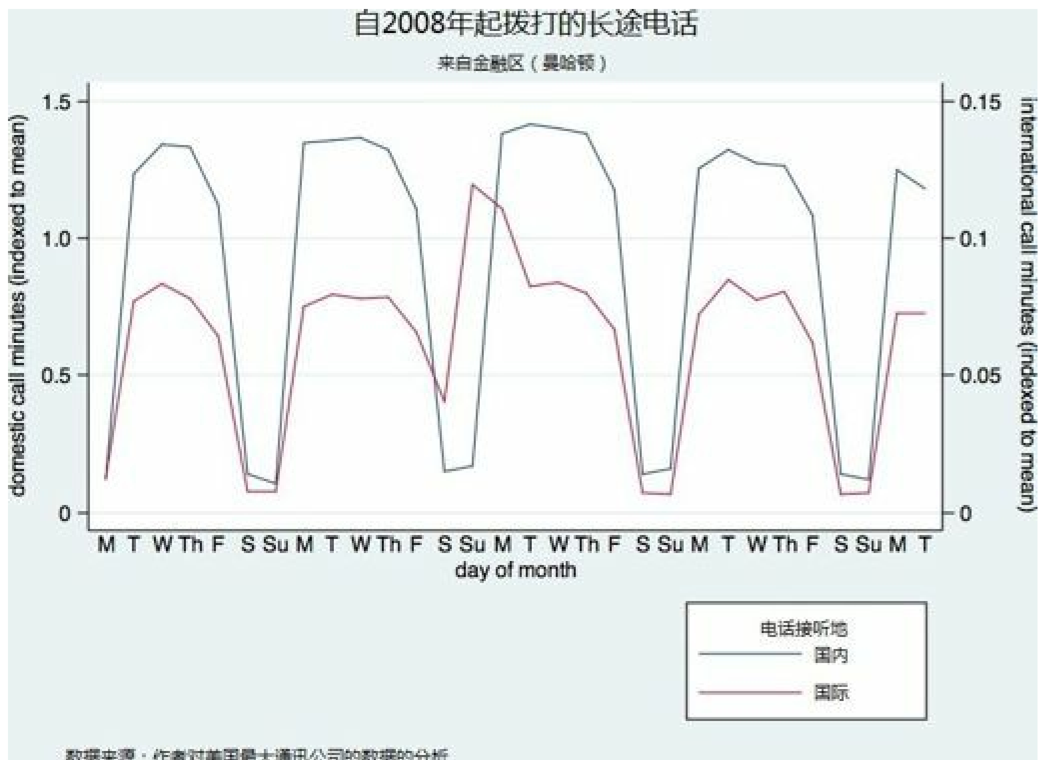
这种利用电信跨越国界的密集交流目的何在？分析每天每小时通话数据的模式，我们能够推断出国际电话的目的。纽约市最“外向”街道的是中央皇后区艾姆赫斯特，我们对比了该区域每天拨向国内和拨向国际的通话量，结果显示一周每天的通话有着惊人的模式。由蓝线表示的国内电话量，每周一和周二会很高，之后会减少。国内电话量在周六会继续下降，而到了星期天会进一步下降。然而由红线表示的国际电话的频率，显示一个相反的效应，工作日期间会有所下降，而在周末会出现峰值，这暗示了电话具有社交目的。工作日和周末之间电话模式的差异表明，来自移民街道的国际电话主要是为了维持家庭和情感的联系。相比之下，在曼哈顿金融区的电话量在工作日会持续的高而周末会很低，而且国内电话比国际电话多。


NYTE
地球脉搏
时区会影响全球通信的节奏。地球脉搏说明了纽约市与255国家一天24小时内国际通话量的变化。有些地区接打电话的量正在减少，而有些地区语音通话量正在增加。国际化城市占了纽约绝大多数的电话往来，图中根据时区的划分做出了相应的明暗效果。

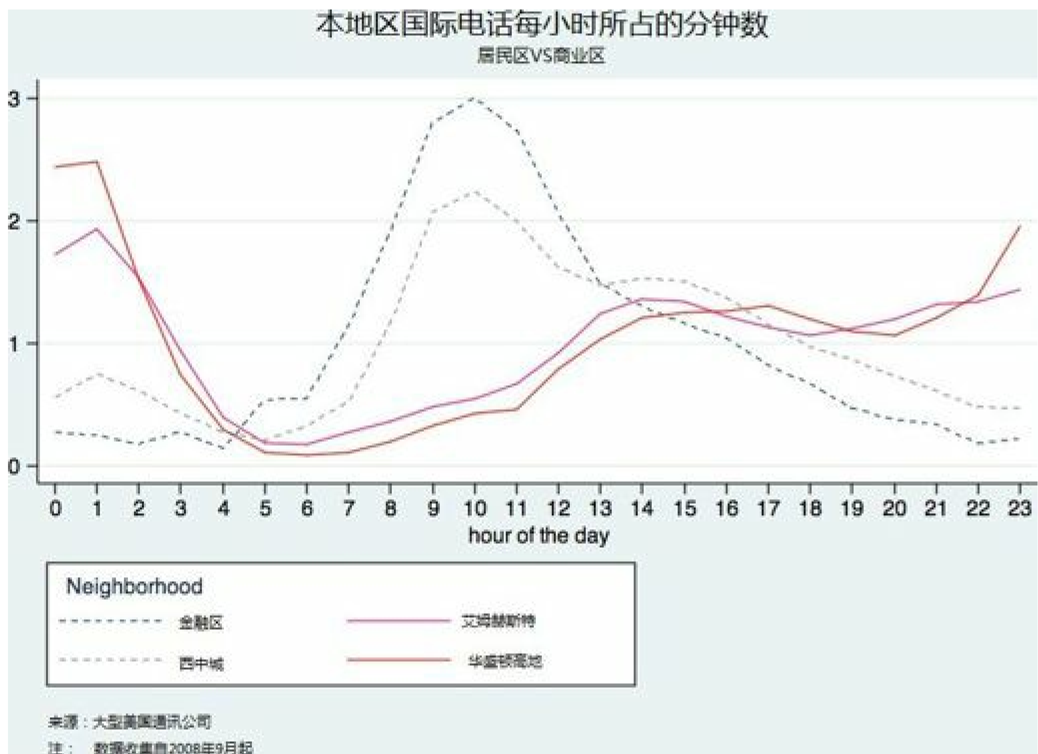
【图6】


自2008年起拨打的长途电话
来自艾姆赫斯特（皇后区）
domestic call minutes (indexed to mean)
international call minutes (indexed to mean)
day of month
电话接听地
国内
国际
数据来源：作者对美国最大通讯公司的数据的分析

【图7】



【图8】



【图9】

比较每天每个小时的通话模式，得到的结论支持这一观点：来自移

民街道的国际电话反应了这些跨国交流的社交属性。图9覆盖了9月份纽约市四个街道每个小时聚合起来的通话模式：以埃尔姆赫斯特和华盛顿高地区（实线）为代表的“外向”移民街道，以及以金融区和中城西区（虚线）为代表的城市商业区。在前期的分析中发现，这些数据与来自线路中心的国际电话时长的平均值挂钩。主要面向商业的街道在白天产生了最高的通话量，而且在早上9点到中午12点之间出现高峰。与此同时，主要的移民街道到晚上非常晚的时候才达到通话的高峰，始终是在晚上11点到凌晨2点。

一位埃尔姆赫斯特地区国家呼叫中心的经理——该地区的人会在电话亭或屋内打国际电话——证实了数据中存在的这些模式：屋内在周末最繁忙，而每周二和周三最清闲。采访结果也支持了这一发现，受访者特别提到，他们在周末电话打的最多也最长，因为那个时候才有时间这样做，而且那个时间接电话的人更有可能在家。打给家里的固定电话是一项重要的策略，因为有一项流行的针对移动电话的无限制国际电话计划，只允许打给固定电话。而且使用预付费电话卡，打给固定电话的费用往往要比打给移动电话的低得多。

尤其对于来自多米尼加共和国的移民来说，“星期天有三件事情：教堂、家庭和休息。”一个上年纪的多米尼亚妇女这样说。她在19世纪60年代后期就来到了纽约市[33]。一般来说，在拉丁美洲的人不会在星期天工作，而且那天大多数的商业场所都会关门。星期天在家打电话给其他家人就能有机会一次性地跟尽可能多的家庭成员通话。去过教堂后，人们聚在家里长时间地共享一顿午餐，而且星期天的电话保证了身在纽约的人与原籍国的家人的联系，通话双方有足够的时间在电话中说说新闻，谈谈管理的业务以及分享一些八卦消息。

结论

虽然在过去的研究中证明，ICT在过去20多年的进步促进了以企业运作为目的的全球化进程，尤其是在像纽约这样的指挥控制中心，但由于

移民的购买力增加，容易获得ICT，像移民这样的因素会如何影响全球化，还有待更深入地探索。这项工作一直想试图通过显示移民确实是塑造纽约全球化的一股重要的力量，以此来揭示移民和电信之间的关系（Castells，1996），而且像电话这样的ICT是推进跨国进程的基本工具，可以帮助我们界定21世纪的移民经验。

本项研究是纽约通过全球电信基础设施与全世界发生联系的一张快照，重点放在了这些无形的全球动态是如何反映了移民区人们的生活。从纽约打向全球的电话比我们现有研究所预期的还要多样化，不仅包括贸易伙伴，还包括城市的一些更重要的移民输出国。各街道所产生的国际电话量的变化强烈地预示了国外出生的人口比例以及中等家庭收入。该方法得到了如下证据的支持：在一个街道范围内，移民原籍国与国际电话目的地国相吻合，在另一个层面上显示了国外出生群体与国际电话量的强关联。最后，移民区的国际电话存在周末的晚上会有很高的通话量这个模式，这反映了电话的目的实际上是为了社交和联络感情。已经证实，城市中一些最具全球化或“外向”的街道拥有着密集的移民人口。在这些地方开展人种学的实地考察，试图回答移民者如何参与全球对话，原因何在。

也许在全球范围内分析无形的信息流动来指导在城市和街道范围的城市规划与政策的实践，这两者的联系会显得很脆弱，但这种信息流概念化后，21世纪移民群体的跨国经验——很大程度是由ICT推进的——是理解美国和其他发达国家城市发展的基础。这方面最重要的事实是，纽约市已经改变了几十年来长期人口下降的趋势，还因为外国移民的影响，使得在过去30年一直保持着人口增长（纽约城市规划部，2004）。由于未来几十年的国际移民，预计城市人口将会继续增长，与移民输出国的通信联系可能会更加密切。

致谢

本文基于作者在麻省理工学院城市研究与规划系的博士论文

（2010）。本文是在完成与麻省理工学院可感城市实验室合作研究后写成的，得到了美国电报电话公司（AT&T）基金的资助。麻省理工学院感知城市实验室研究人员弗朗西斯科·卡拉布雷斯（Francesco Calabrese）和乔恩·里德斯（Jon Reades），他们创建了本研究使用的聚合电话记录的数据库，乔恩开发了一种可以映射电话量的地理空间方法。特别感谢弗朗西斯科和乔恩对本项工作所作出的贡献，以及作者所在的博士生委员会——弗兰克·利维（Frank Levy），曼纽尔·卡斯特（Manuel Castells）和卡洛·拉蒂（Carlo Ratti）教授——感谢他们的指导。

参考文献

1．Castells, Manuel. 1996. The Rise of the Network Society. Oxford: Blackwell.

2．Derudder, Ben. 2008. Mapping Global Urban Networks: A Decade of Empirical World Cities Research. Geography Compass 2(2): pp.559-575.

3．Federal Communications Commission. 2008. Trends in Telephone Service. Washington D. C.: Federal Communications Commission. August.

4．Federal Communications Commission. 2009. 2007 International Telecommunications Data. Washington D. C.: Federal Communications Commission. June.

5．Gottmann, Jean. 1967 (1961). Megalopolis: the Urbanized Northeastern Seaboard of the United States. Cambridge, MA: The MIT Press.

6．International Monetary Fund. 2008. Direction of Trade Statistics. Washington D. C.: The Fund Levitt, Peggy and Nina Glick-Schiller. 2004. Transnational perspectives on migration: conceptualizing simultaneity. International Migration Review 38(3) pp.1002-1039.

7．Massey, Doreen. 1994. Space, Place and Gender. Minneapolis: University of Minnesota Press New York City.

8．Department of City Planning. 2004. The Newest New Yorkers 2000. New York: Department of City Planning, Population Division.

9．New York City Department of City Planning. 2009. Current Population Estimates.

10．http://www.nyc.gov/html/dcp/html/census/popcur.shtml (letzter Zugriff: 5. July 2013)

11．Pool, Ithiel de Sola. 1983. Forecasting the Telephone: A Retrospective Technology Assessment. New Jersey: Ablex Publishing Corporation.

12．Robinson, J. 2002. Global and world cities: a view from off the map. International Journal of Urban and Regional Research 26(3) pp.531-554.

13．Rojas, Francisca M., Clelia Caldesi Valeri, Kristian Kloeckl, and Carlo Ratti, ed. 2008. NYTE: New York Talk Exchange. Cambridge, MA: SA+P Press.

14．Sanjek, Roger. 2000. The Future of Us All: Race and Neighborhood Politics in New York City. Ithaca, New York: Cornell University Press.

15．Sassen, Saskia 2013 (1991). The Global City: New York, London, Tokyo. Princeton, NJ: Princeton University Press.

16．Sassen, Saskia, ed. 2002. Global Networks, Linked Cities. New York: Routledge.

17．Vertovec, Steven. 2009. Transnationalism. London and New York: Routledge.

注释

［1］2008年，丹麦艺术家奥拉维尔·埃利亚松（Olafur Eliasson）在纽约建造了四个人造瀑布。人造瀑布由巨大的水泵将水从37米的空中倾泻而下。根据建造前的预估，整个装置将耗资约1,500万美元，并为纽约带来约5,500万美元的旅游收入。（译注）

［2］原文为：vantage points of the，原文缺漏，译者根据文意把句子补充完整。（译注）

[3]“现实挖掘”（reality mining）是随着数字足迹研究的发展而出现的一种研究方法，主要是通过手机、GPS、可穿戴设备等移动设备来采集人们到哪里去、停留时间长短、和什么人有什么样的互动等等之类的数据，并对其进行分析和展示。传统的、基于互联网数据的客户行为研究，一般是基于客户在某一个网站上的行为和言论而进行的，研究者只能了解客户在某一方面的行为；而现实挖掘方法使用的新型技术设备，能够全面采集客户在各个方面的行为，从而能够得到更加完整的用户画像。与一般的数据挖掘方法相比，现实挖掘更加强调将数据分析的能力与社会学等相关学科的理论和实践融合在一起。（译注）

[4]文中“自上而下的、老大哥式监控”（large-scale top-down big brother threat on privacy issues），可参见乔治·奥威尔的作品《1984》。“小范围内人们自下而上的、小女孩般的观察”（more local bottom-up little sister types of people monitoring）中“自下而上的权力”可参见法国哲学家米歇尔·福柯的著作中“微观权力论”相关内容。传统理论一般认为，权力会表现出三个要点：权力被占有，通过镇压（压抑）实施，自上而下。而福柯则认为，权力首先并不是一种所有权，它是被行使而不是被占有；其次，权力不是单纯的压抑，它也具有传播、训练、塑造及至生产的功能；第三，权力不仅是自上而下的单向过程，不是集中于某些机构或阶级，而是有无数的作用点。（译注）

[5]Open311: http://open311.org

[6]开放数据目录：http://oad.simmons.edu/oadwiki/Data_repositories

[7]OpenStreetMap: www.openstreetmap.org

[8]纽约时报在2011年3月5日的报道：http://www.nytimes.com/2011/03/05/science/05legal.html

[9]苹果公司iPhone数据的追踪：http://radar.oreily.com/2011/04/apple-location-tracking.html

[10]苹果公司几周后更新了它的操作系统，使得在当时建立类似crowdflow. net数据库变得不再可能。

[11]参见案例1: 12-cv-00784-LAK, Aka Carniol v. Yassky等人。

[12]Moovel: http://www.moovel.com

[13]对这项研究而言，“实时城市技术”和“实时技术”定义是提供动态数据和实时或接近实时监测的情境技术。例子包括交通拥堵报告，GPS预测公交到达时间还有空气质量监测。有

关“开放政府”的数据并不包括这项研究中因为它并不是“实时”可得的。

[14] Golder and Macy

[15] Williams, Robles, and Dourish, 2008.

[16] Pike Reasearch, 2010.

[17] Transit Watch, 2000.

[18] Transit Watch, 2000.

[19] NCB, 1992.

[20] You, 1991.

[21] Infocomm Development Agency, 2005.

[22] Department of Statistics Singapore, 2012

[23] 相对照的是，美国每100人拥有90.78部移动电话，香港每100人拥有179.39部移动电话。

[24] Department of Statistics Singapore, 2012

[25] Ito, et al, 2005

[26] "The full ST interview …", 2011

[27] 本项研究所涉及到的电信提供商在纽约市的国际电信市场所占的份额不能公开，但是根据美国联邦通信委员会（FCC）2007年给出的国际电信数据报告，2004年该公司所产生的计费国际电话时长超过了美国总数的1/4。

[28] 此处使用呼叫时长而不是呼叫次数，来作为标准的分析基础。这是由美国联邦通信委员会和美国电信市场调研公司（TeleGeography）基于他们的测量和分析得出的协议。

[29] 墨西哥被认为是北美，不属于中美洲和加勒比地区。

[30] 图3所表示的回归并不包含所有223个国家的通话数据，因为2006－2008年美国社区普查没有列出原籍国的数据，而这些数据跟纽约街道呼叫的目的地国的数量一样多。不过，有69个国家的关于每个区的通话时长和国外出生人口的数据表示了出来。

[31] 特里贝克地区／唐人街既不包含在此表中，也没有做进一步分析，因为这个节点很高的呼叫量不是代表其各自街道的环境，但是代表纽约全市电信基础设施的一个功能。这个线路中心至少包含了两个主要的长途交换站，电信提供商以此路由了城市大部分长途通信量：托马

斯街33号和美洲大道32号。这些交换设备收集和路由了超出它们附近区域的长途通信量，以此可以说明一个情况，即利用线路中心地理边界内的人口可以解释哪些地方可以看到更多的电话聚合。

［32］为了将纽约市的收入分成低、中、高三个档次，我根据由美国大型电信提供商提供的线路中心数据得出了纽约中等家庭收入，即47,900美元。我设定等于或低于中等家庭收入80％的定为低等收入（≤38,320美元），间于80％至150％的定为中等收入（38,320美元至71,850美元），等于或高于150％的为高等收入（≥71,850美元）。该方法源自2006年布扎（Booza）等人的研究。特里贝克地区／唐人街的线路中心没有用到这种回归分析，因为它不能代表其地理附近的信息；肯尼迪机场也排除在外，因为它没有相应的人口统计信息；而曼哈顿一个未命名的线路中心因为没有相应的通话数据也排除在外。还有其他国际电话的异常值高出均值两个标准差，这些数据也同样会被删除。

［33］由作者翻译自西班牙语。

第二章　表现——模型与可视化

作为数字公共空间的城市——关于设计实时城市数据平台的几条意见

作者　克里斯蒂安·克洛克（Kristian Klockl）

译者　党霄羽

设想一下在街上遇到一位朋友。你们两位都停下脚步，交换信息——诸如相互问候，讲讲自己的趣事，谈谈天气，等等。在进行谈话的时候，你根据你对这个话题的了解以及你对你朋友的背景的了解来选择自己的措辞。当他皱眉的时候，你知道他表示反对；他的身体语言让你明白他想要插话，想要应答。于是你打住话头，给他发言的机会。你注意到你的朋友在看表，他变得不耐烦起来。他发现你注意到他的焦躁，解释说他要去赴约。这些都是通过许多微信号的交换实时进行的。鉴于信号交换是可感的，信号强度足够在大脑中处于活跃状态，能够被接受者所理解，因而我们能够有效捕捉到这些交流的动态。

我们也是以同样方式应对环境信息（environmental information）的。然而，这种有趣的现象随着距离的增加会变得困难，或者这些信号无法被人类感知，无法进入大脑。现在，假设我们可以读取所有来源的环境信息，基于这些信息作出决策，就好像你和朋友间的对话那样是一场有效的交流。

现在的城市空间中架设着数字网络和数字系统，创造着反映人类活动的信息。尽管大部分利用电子化手段进行管理的城市系统围绕着自己的目的选择性地产生运营数据，这些系统并不为其他系统或者公众提供直接的入口来分享这些数据。于是，数据信息在某个特殊领域的范围内，反映出城市中不同地点的人类活动。在“实时新加坡！”（LIVE Singapore!）的项目中，我们开放了一个公开的平台，用来搜集、整合、分配大量这样的实时数据资源，鼓励开发者们把这些数据流用于创

造有意义、有价值的城市工具。

“实时新加坡！”承认，城市系统运行正在产生海量数据作为其副产品，而这些数据中包含的有意义的信息，已经超出了原有系统的界限。在这个意义上，“实时新加坡！”并不是要尝试解决城市的具体问题，而是对新的数字工具进行了探索，这些新工具产生了一种处理城市问题的全新方法。

新加坡：一座理想的实时之城？

新加坡在各种意义上都是探索实验一个开放的实时平台的理想地点，其中的原因是多种多样的。新加坡是一座城市国家，其全部领土只有一座城市，在动力学（dynamics）意义和政治统治意义上都是如此。新加坡有5,076,700位居民（基于2010年的调查数据），他们居住的这片土地仅有710.2平方公里。新加坡的领土是一座小岛，仅仅通过两座桥和马来西亚这块陆地相连，供摩托车往来交通。通过海运或者空运，就能满足其他的载人或者载客的需要。樟宜国际机场是东南亚地区的交通枢纽。新加坡的集装箱码头是世界第二大的集装箱码头，同时也是世界第一大集装箱转运港口。新加坡是一个高度发达的城市，有先进的基础设施信息网络，还有能娴熟使用高科技设备的人群。新加坡的手机市场渗透率在2010年达到了143.6％（数据来源：IDA——新加坡资讯通信发展局）。此外，这是一个高速发展的城市，因而新加坡人对城市结构的巨大变化见惯不惊。自1965年新加坡独立以来，这个国家就经历了经济和基础设施的双重高速发展。新加坡的地铁在1987年才开放，现在已经扩展到了四条线路，共有87站，里程达130千米（数据来源：新加坡陆路交通管理局）。

由于这些原因，新加坡为实验城市实时数据平台提供了得天独厚的条件，能够调查与这个城市相关的所有研究和开发。基于这些特殊的条件，我们的研究所面临的最大风险，就是并不是所有的论证结果都适用于其他城市。

开放的数据计划与数据平台

“实时新加坡！”受到了最近的开放数据计划的启发。在2009年5月，美国政府执行了数据政府（data. gov, 2011）计划的第一期，公开了那些不敏感的历史数据。“其目的是，让搜索、下载、使用联邦政府产生和持有的数据集变得轻松。”这些数据涵盖了经济发展、自然科学、教育改革、推动市民参与、政府经费削减等领域，项目的目的之一是“将属于人民的数据还之与人民”。这项计划被许多国家和城市争相效仿（例如英国伦敦、旧金山），通过这些平台提交了许多份申请要求获得这些公开数据。

除了公共数据之外，像Pachube（Pachube, 2011）这样的平台使得人们能够分享传感器数据（“数据经纪”），这是一种物联网的基础设施。

开放平台项目

“实时新加坡！”在上面描述的这些计划的基础上建立起来。这是一个灵活而具备可扩展性的开放平台，为搜集、整合、分配城市的实时数据流创造条件。这个项目的自我定位是一个向开发者社区授权的平台，开发者可以通过平台创建使用数据服务集合的申请。正如我们相信数据政府项目所表现出来的创新性一样，我们非常相信城市以及城市居民的创造性潜力，让城市的居民、规划者、公司和政府使用实时的数据流，为新兴经济体建立起从数据中取得价值的新方法的基础。

原先的实时数据项目，即使数据来源比较单一，也已经产生了一些有趣的结果。愈加明显的是，不同数据之间的整合可以带来更加丰硕的成果。这个项目所使用的数据可以分为三组：第一，作为现存网络的副产品的数据；第二，从标签或者传感器中搜集到的数据；第三，由人群主动分享产生的数据。

同样，数据提供方能够给数据接入者提供几乎没有经过准备和分类的数据，但又不会降低数据介入者的参与积极性；同时，又能够让数据流尽可能地具有可互操作性和连贯性，这两者之间的冲突能够通过这一

平台来解决。在这个意义上，“实时新加坡！”增加了城市系统的多样性，因为“计算机互联会带来多样性。由于计算机连接或者互联的激增，多样性就会增加，相应的，信息量也会增加。”（Taylor, 2001, p.139）。这就对平台提出了另外一个要求，那就是为了吸引活跃的大型开发者社区，必须提供容易上手的接入方式和可编程序。

用户们自己创造了不同类型数据之间富有意义的联系

传统上认为，在线平台的主要价值就是这些平台积累的数据能够使用户行为和用户偏好得到深入分析。“实时新加坡！”的做法与之有根本差异。它处理的是数据流，因而并不存储任何数据。总是由数据提供方来决定哪些数据是可见的，这简化了数据的所有权、隐私性与安全性的问题。因而这一平台能够提供一致的数据流，而又不对之加以储存。历史数据是存储在数据提供方那里的，当提供方设备与平台相连接时，平台就可使用这些数据。

因而，“实时新加坡！”的主要价值在于连续记录提供方提供的实时数据流和用户创建的申请。这还带来了一个关于这种联系的语义学问题：不同类型的数据流之间的联系的意义和价值是什么？我们的研究方法是，让用户进行这些联系，并且将之描述出来，这样就能逐渐发展出一套能够说明平台进化的还是管理系统。

反馈回路（**feedback loop**）

城市中的人类活动在不同的技术系统之中产生数据。例如，搭乘地铁就要刷一下电子车票才能进出站，这个时候包含一系列信息的电子轨迹就已经生成，包括车票编号、车程的起止时间，坐的是几号线。把这些信息加在一起，数据流就能够显示出在任何特定时间列车上的总人数，列车在哪里晚点会给多数乘客造成影响，等等。通过将这些数据交还给公民，他们就可以决定如何更有效率地使用这些系统。正如神经控制论的前沿人物杰伊·福里斯特（Jay Forrester）的解释，这就创造了一个反馈回路：“在一个信息反馈系统中，条件被转化为信息，这是进行

控制行为，并改变原有条件的基础。这个过程是连续的。我们没法说清这个循环在哪里开始，在哪里结束。这是一个封闭的回路。”（Forrester，1961，p.61）“现在可以得到的关于过去的信息，正在被用作未来决策的基础。”在这样一个系统中，情况当然还是如此。（Forrester，1961，p.61）现在，像上面所说的那样处理实时数据让过去和现在贴得更近了。福里斯特指出，这样一个系统中的延迟反而可能具有稳定系统的作用，这种延迟的缩短可以增加灵活性，帮助我们以一种更有效的方式来利用资源和能量。

这提出了一些重要的问题：实时反馈回路会不会让系统的稳定性变差，或者，这个系统能稳定吗？我们或许从唐娜拉·梅多斯（Donella Meadows）那里得到一些启示。她指出，系统各部分之间的内在联系起到了特殊作用——“改变系统的内在联系，系统就会发生极大变化”（Meadows, 2008, p.17）。

系统图像

在处理城市网络系统的实时数据的时候，可感城市实验室不仅仅把重点放在平台的软件开发与分析上，而且还强调对城市语境中根植于数据收集和研究的有效内容进行交流。数据可视化（Data Visualizations）已经被证明，在这项任务中特别有效。正如梅多斯所说：“词语和句子必须是一个接一个按逻辑顺序线性排列的。而系统是同时出现的。［……］为了讨论的需要，讨论话题所使用的语言和话题所讨论的现象必须有共同的特点。图像比词语更适合充当这种语言，因为你在一瞬间就能看到图像的所有内容。”（Meadows，2008，p.5）能够一瞬间呈现出图像的所有内容，就是因为图像的这个优点，控制室所处理的是大量数据，呈现出来的是复杂的图表，而不是简单地给出数字信息。数据可视化的第二个优点，是这会为观测者带来额外的运算能力。对观测进行模拟显示的可视化城市实时数据需要调动观测者的理解力。在观察的过程中，她会将自己以前的知识和经验带入到数据的分析和过滤之中，这就

好像一个人在决定走哪条路之前，站在高楼大厦之巅，俯瞰交通状况一样。在这种情况下，观测者把她过去的经验带入进来，把这些和她实时观测到的情况综合在一起。

“实时新加坡！”城市样本



【图1】

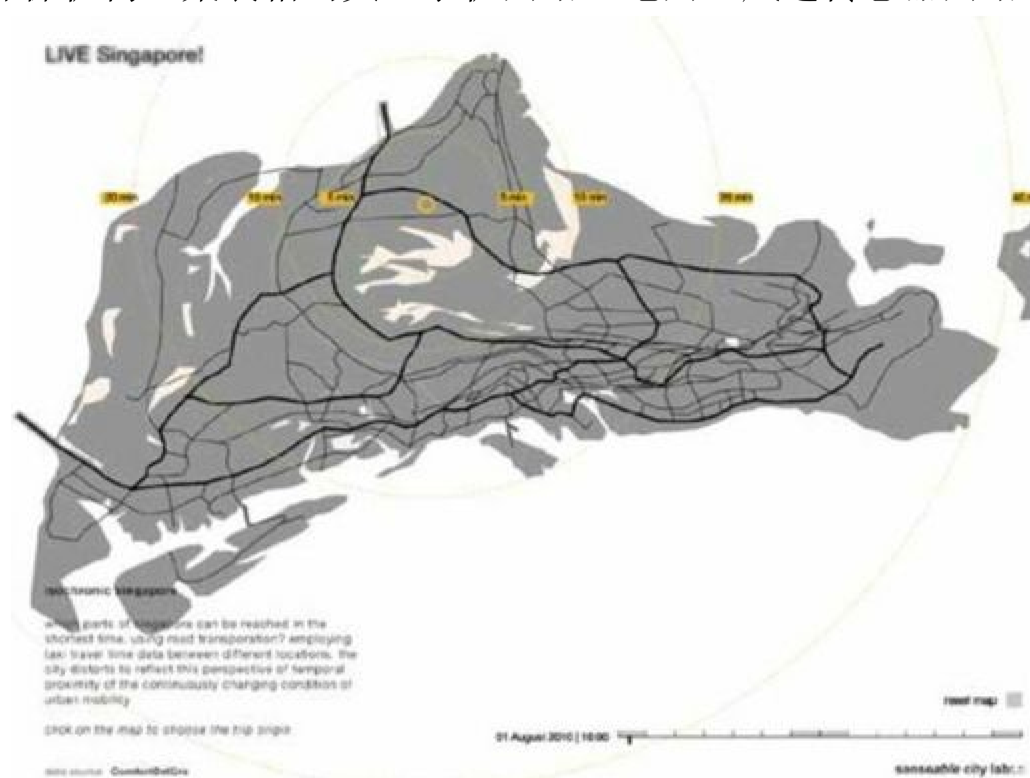
“实时新加坡！”着力于让市民能够获得城市实时数据，在这种精神的指引下，通过几个重复的步骤，这一原型已经得到了发展，并和城市中的公众分享，以激发公众的讨论，讨论处理这种数据对于规划、管理城市以及城市的生活有何种潜力。

为了达到这些目的，首批的两个样本应运而生：2011年4月在新加坡艺术博物馆（Singapore Art Museum）举办的“实时新加坡！展览”以及2012年1月在新加坡国立大学亮相的，由SMART（Singapore MIT Alliance for Research and Technology）组织的未来城市移动性研讨会上所展示的“对城市移动性的可视化探索”项目。

第一个城市样本：“实时新加坡！展览”

“实时新加坡！”展览在艺术博物馆（博物馆前身在19世纪中叶是圣

约瑟夫学院）举行，展览中使用了五个投影机和一个LCD屏幕，显示着动态的、半交互式的可视化数据。这些数据来自结合的数据集和实时数据流，都是由重要的城市网络提供的。这些城市网络包括：机场、出租车、环保机构、集装箱码头、手机网络、电网、风速传感器网络。



【图2】实时（Isochronic）新加坡——这个城市的最大的出租车运营商（有16,000辆出租车）使用的GPS定位和测速数据。这些数据被用作路面行驶时间的总指标。随着交通情况的好坏，我们在新加坡行进的速度时快时慢。从你家到目的地需要多久呢？在这张实时地图上，出行所要耗费的时间在不断变化——这种地图还显示出平日和周末的不同。

第二个城市样本：城市移动性的可视化探索

在对新加坡艺术博物馆的展出进行了学习之后，联合了新加坡路面运输司（Land transport authority），我们开发出三款交互式的应用，对新加坡的运输基础设施产生的数据所提供的丰富信息进行了分析。这些应用的原型曾经在2012年新加坡的“城市移动性的可视化探索”中展出。

其中一款应用叫做接触公交（TOUCHING BUS RIDES），是由“城

市移动性的可视化探索”项目开发的，致力于发展城市公交网络。新加坡的公共交通系统需要乘客在上下公交或者地铁的时候刷卡。这些行为转化为各种各样的基于乘车距离的合理票价，同时也能在任何时间提供有关乘客上车的信息。

交互式的接触界面让用户能够主动探索新加坡的公交网络，查看大部分乘客都在哪里上下车，如何换乘，以及这些模式在一天之内的变化。

用户可以在不同的可视化界面之间进行切换，以取得同一组数据的不同视觉效果。选中的公交线路、突出显示的公交车站以及在应用时所选择的停留时间，在所有模式下都是一致的，而且应用可以无缝在各种不同的虚拟的可视化图形之间切换。这款应用展示了各种汽车时间表，上面有标示说明乘客在每一站的上下车情况，还能够跨时段进行比较。用户还能够在新加坡地图上选中任何一条公交线路。这个可视化图形的时间轴还标示出了能够展示出乘客数量的直方图。用户可以迅速浏览到这些数据，从每隔30分钟变换一次，到每隔一天变换一次。

对城市样本的观察

上面描述的两个城市的样本旨在说明城市空间中产生的实时数据十分丰富，很有价值。在我们的研究的语境中，这些城市样本有多种作用：可以把不同领域的关于同一原型的不同应用结合起来，这些原型能够传达这个项目的理念；可以让公民使用这个项目和原型进行试验，为进一步的开发提供有用的反馈；可以通过让这个项目成为现实，来让项目外的参与者看到这座丰碑。

创造这些城市的样本包括为数据分享协议打下基础，这一协议是在城市系统的操作者在整个过程中和他们公司的各部门之间的合作之中达成的。我们的研究通常都不关注这些工作。这一过程的价值在于，让我们能够了解和城市数据有关的挑战与机遇。

打开数据仓库

因为这些系统都不是为了分享这些产生的数据而设计的，所以为了有意义地使用和分享这些数据，有必要对这些系统进行了解。在通常情况下，这会在技术上带来巨大的挑战——这意味着要么无法在系统外部使用这些数据，要么就要进行为数不小的投资。此外，数据的内部操作也是一个关键因素，这不仅是因为涉及到不同机构和系统之间的数据分享，还涉及在同一个机构内部使用不同的系统。在一个组织中将不同系统产生的数据整合在一起，其复杂性堪与整合不同组织之间的数据相比。

这些数据的使用已经超出了它们原有的目的[1]，这就需要深入了解这些数据如何产生、在何种语境下产生。例如，如果一个传感器显示了大楼外面温度读数，那么就需要知道它的位置是不是接近通风口，是在向阳处还是背阴处的。像“实时新加坡！”这样的项目驱使着数据的提供方对机构所产生的数据形态进行深入挖掘，在组织内部充分挖掘数据的潜力。系统数据通常是归IT部门来管，这些数据和规划与运行部门之间存在着一定的物理距离。其结果是，那些负责掌管数据系统的人侧重于从技术的角度让系统操作不出差错，然而那些负责规划和运行的部门几乎无法接触到数据中隐藏的信息。

促进新城市方案的理念之间的共同发展

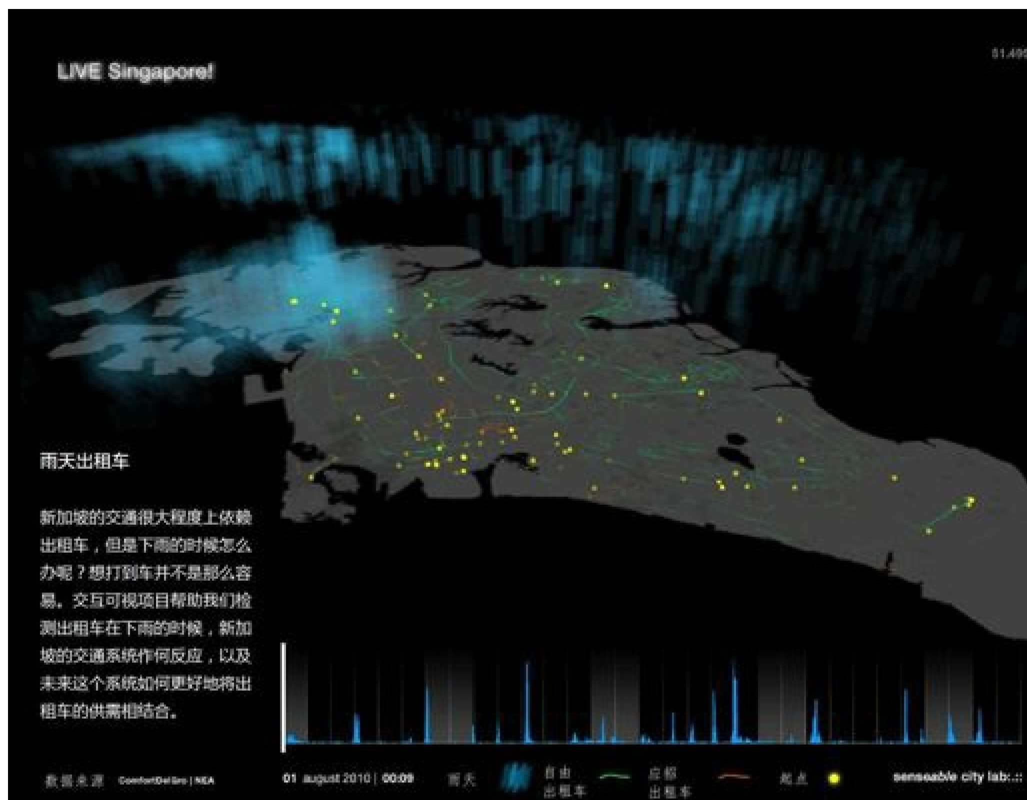
开发一个有意义的城市系统数据呈现方式，为人们接触隐藏信息创造便利是一个关键因素。这意味着把不同学科的相关人员以及人才聚集到一起，来开辟一块新的研究范围进行创造性的思考。这种做法的一个例子就是ChangiNow计划，这一计划就从上面所说的城市样本中脱胎而来。一旦我们想办法去处理从出租车公司以及机场得来的数据，就诞生了一个想法：把飞机场关于实际到达旅客数目的数据和空闲出租车数量相结合，机场的出租车的供需就可以更好地匹配。这个关于交通衔接的例子是常常被忽视的。

再举一个例子。“接触公交”的触摸式列表把各类相关人员聚集在了

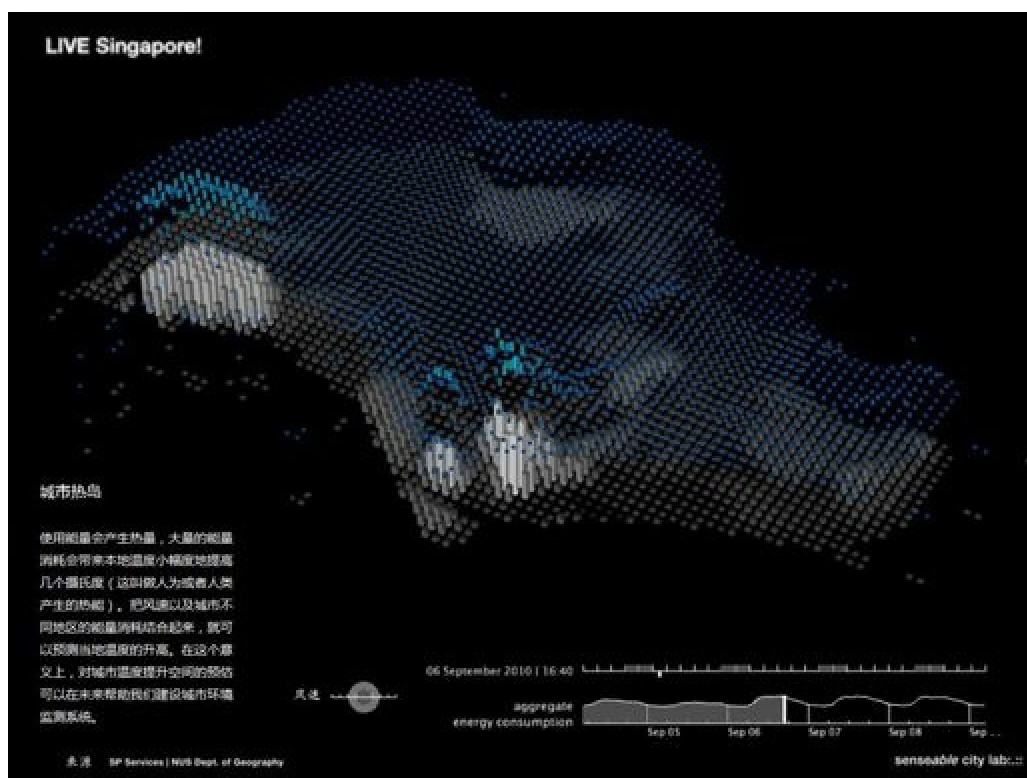
一起。他们的讨论导致了三种应用：第一，创造了动态的公交路线，这一路线随着在拥挤的交通中滞留的乘客数不断变动；第二，建立了动态的特快公交路线，只要实时需要的模式显示这种需求，就会在遥远的车站之间建立直接的联系；最后，在公交站安装动态信息展板，告知正在等车的人们下一辆公车的位置，以及它的到站时间。

城市系统数据的跨学科的工作方式，反映出数据作为“边界对象”的作用，因为“（这些对象）在不同的社会有不同的意义，但是它们的结构之间的共同之处，让它们能够在不同的社会得到辨认和转译。”（Star & James, 1989）举个例子，在电话网络中产生的数据能够让我们得到关于系统不同方面的情况，诸如现存的基于邮件群组的社交网络、空间移动状态、经济行为等等。在这种情况下，通过交互式的视觉界面和跨学科的方式来处理城市数据，能够关于城市动力学形成新的问题，这超越了仅仅选择一个最佳解决方案。当这些年进行这些研究和开发的时候，这种处理方式就变成了现实。创新带来的重要的挑战是，如何界定这些数据的所有权，平台如何简化咨询和交流方式，从而在数字领域重新提出城市公共维度的问题。

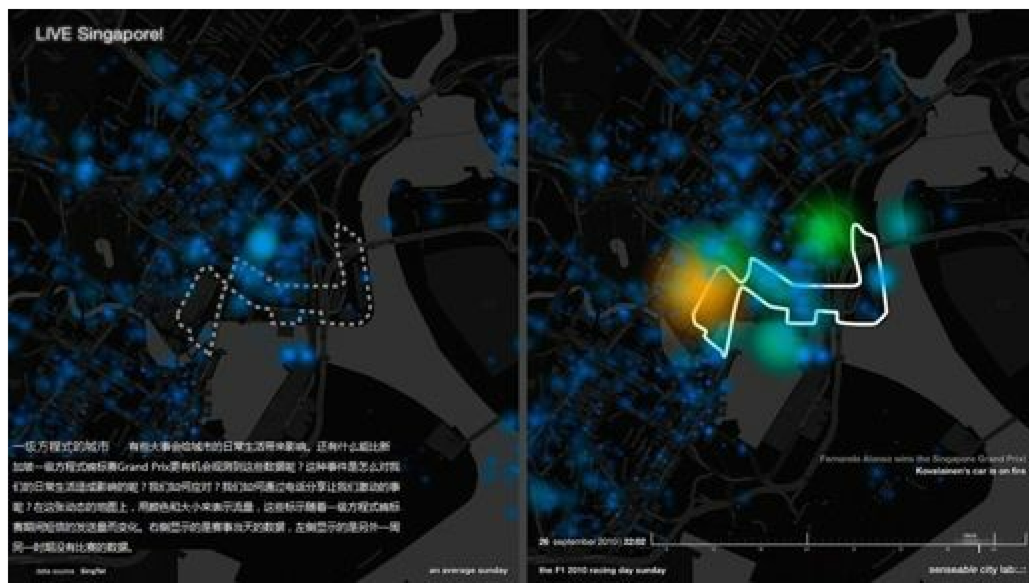
图示


LIVE Singapore!
雨天出租车
新加坡的交通很大程度上依赖出租车，但是下雨的时候怎么办呢？想打到车并不是那么容易。交互可视项目帮助我们检测出租车在下雨的时候，新加坡的交通系统作何反应，以及未来这个系统如何更好地将出租车的供需相结合。
数据来源 ComfortDelGro | NEA
01 august 2010 | 00:09
自由
出租车
起点
senseable city lab:::

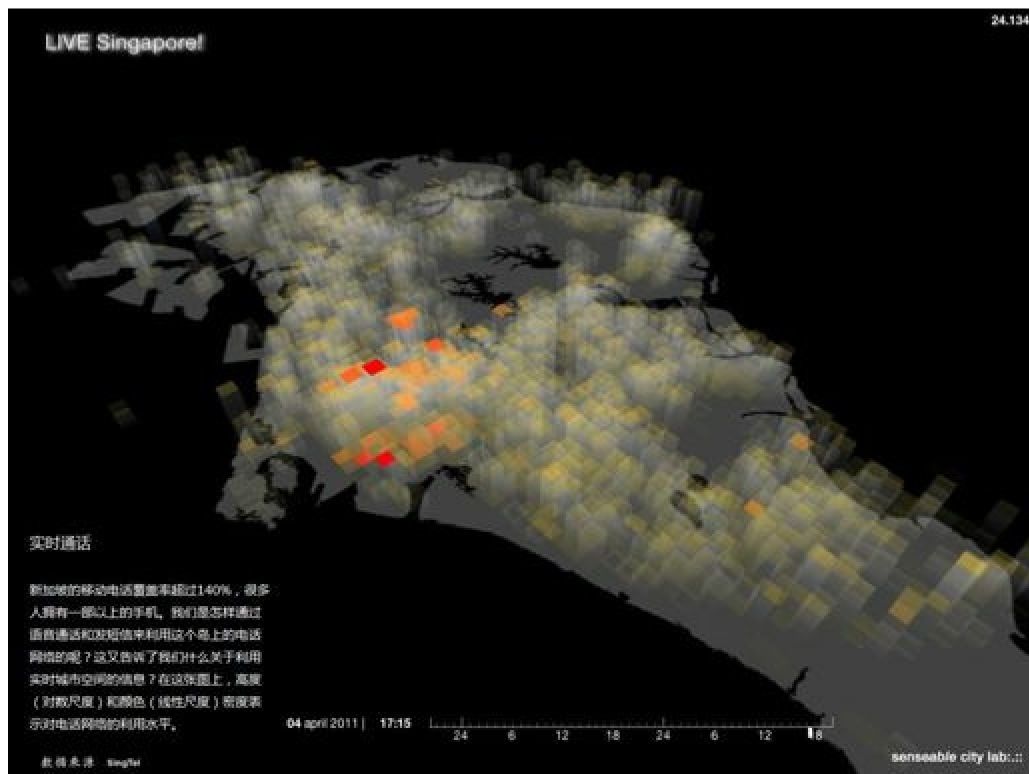
【图3】


LIVE Singapore!
城市热岛
使用能量会产生热量，大量的能量消耗会带来本地温度小幅度地提高几个摄氏度（这叫做人为或者人类产生的热能）。把风速以及城市不同地区的能量消耗结合起来，就可以预测当地温度的升高。在这个意义上，对城市温度提升空间的预估可以在未来帮助我们建设城市环境监测系统。
风速
06 September 2010 | 16:40
aggregate
energy consumption
Sep 05
Sep 06
Sep 07
Sep 08
来源 SP Services | NUS Dept. of Geography
senseable city lab:::

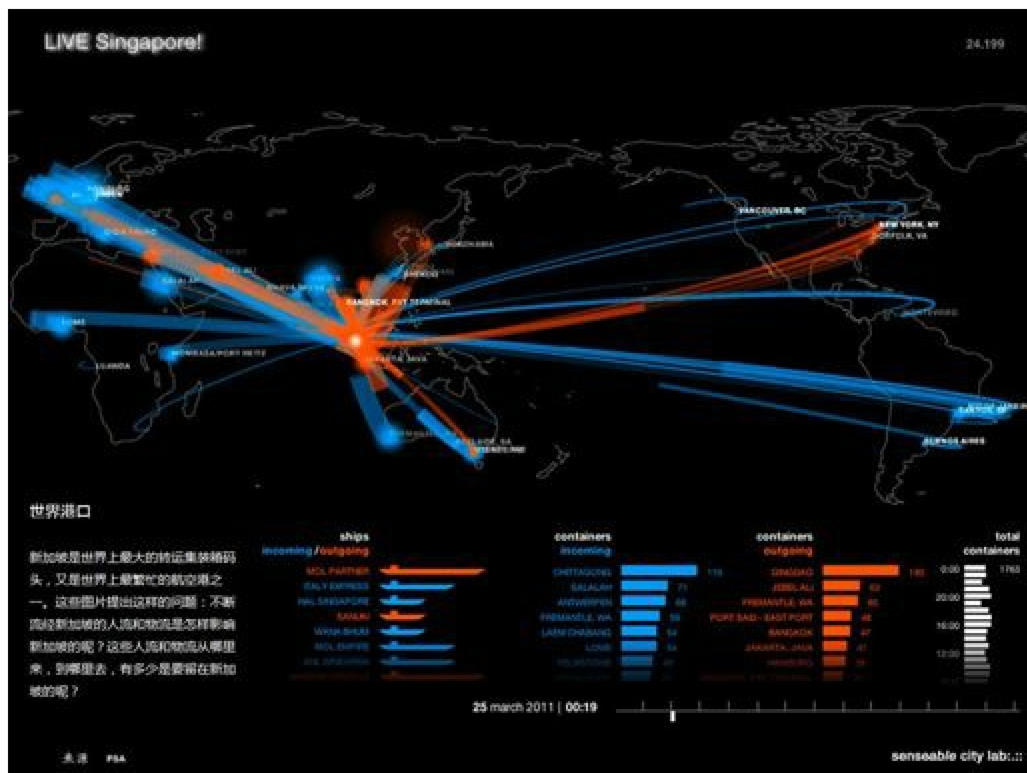
【图4】


LIVE Singapore!
一级方程式的城市
Fernando Alonso wins the Singapore Grand Prix
Kovalainen's car is on fire
26 september 2010 | 22:02
an average sunday
the F1 2010 racing day sunday
senseable city lab

【图5】


LIVE Singapore!
24.134
实时通话
新加坡的移动电话覆盖率超过140%，很多人拥有一部以上的手机。我们是怎样通过语音通话和发短信来利用这个岛上的电话网络的呢？这又告诉了我们什么关于利用实时城市空间的信息？在这张图上，高度（对数尺度）和颜色（线性尺度）帮我表示对电话网络的利用水平。
04 april 2011 | 17:15
24 6 12 18 24 6 12 18
数据来源 SingTel
senseable city lab

【图6】


LIVE Singapore!
24.199
世界港口
新加坡是世界上最大的转运集装箱码头，又是世界上最繁忙的航空港之一。这些图片提出这样的问题：不断流经新加坡的人流和物流是怎样影响新加坡的呢？这些人流和物流从哪里来，到哪里去，有多少是要留在新加坡的呢？
ships
incoming/outgoing
containers
incoming
containers
outgoing
total
containers
25 march 2011 | 00:19
PSA
senseable city lab:.::

【图7】


LIVE Singapore!
20 min
10 min
5 min
5 min
10 min
20 min
共时新加坡
新加坡的哪个区域可以在最短时间内通过陆路到达？利用在不同地点之间的打车时间的数据，城市会改变形状，以反映当前随着城市交通状况不断变化的最短距离。
reset map
01 August 2010 | 10:00
data source ComfortDelGro
senseable city lab:.::

【图8】

参考文献

1．Forrester, JayWright. 1961.Industrial Dynamics. Vol.2. The MIT Press Cambridge, MA.

2．Meadows, DonellaH, and DianaWright. 2010. Die Grenzen des Denkens: wie wir sie mit System erkennen und uberwinden konnen. Munchen: Oekom-Verlag.

3．Murty, Rohan, Abhimanyu Gosain, Matthew Tierney, Andrew Brody, Amal Fahad, Josh Bers, and MattWelsh. 2008."CitySense: A Vision for an Urban-scale Wireless Networking Testbed." In Proceedings of the 2008 IEEE International Conference on Technologies for Homeland Security, Waltham, MA.

4．Paulsen, H. ,and U. Riegger. 2006. "Sensor GIS-Geodaten in Echtzeit." GIS-Business8(2006):17-19.

5．Star, Susan Leigh, and James R. Griesemer. 1989. "Institutional Ecology,`Translations' and Boundary Objects: Amateurs and Professional sin Berkeley's Museum of Vertebrate Zoology, 1907-39. "Social Studies of Science 19(3) (August1):387-420.

6．Taylor, MarkC. 2002. The Moment of Complexity: Emerging Network Culture. University of Chicago Press.

7．Zardini, Mirko, and Wolfgang Schivel busch. 2005.Sense of the City: An Alternate Approach to Urbanism. LarsMullerPublishers.

城市的画像与漫像

作者　佩德罗·克鲁茨（Pedro　　Cruz），彭索尔·马沙多（Penousal Machado）

译者　党霄羽

导论

源自城市的丰富数据可以让我们了解人们怎样组织起来，无论组织的规模有多大。可视化城市话语提供了种种方法，有抽象的，也有直观的、形象的。在本章中，我们使用一种不那么抽象、不那么简单的方式来处理可视化城市问题。我们超越了科学的可视化的处理方式，转而投向我们称为具体的可视化方法。通过这种方式，我们退一步来审视可视化必须以抽象直接的方式呈现数据这种理念。通过一种形象的方法，我们把图像与漫像引入到了数据的国度，使用可视化的隐喻，引入变形的视觉效果，来强调数据的某些特定方面。这使得我们能够使用有趣的类比，构建视觉效果，来向学术圈外的一般大众传达城市的本质特征。

这篇文章将描述了两个可视化的案例，既有抽象方法的案例，也有具体方法的案例，但是没有涉及具体的技术细节，而是试图用一种可信的话语，使用图像与漫像的方式，在可视化数据的语境中，把这些呈现出来。

可视化的具体方法

信息可视化是一个跨学科的领域，涉及平面设计、人机交互、计算机图像学以及数据挖掘。这个学科的目的是为广大受众综合大量的数据，从数据中提炼并澄清信息。这里所说的具体方法，分为相片、图像、漫像以及扭曲等方法，从而得以描述可视化的不同方式。这种分类学背后的概念是著作权（译注：根据上下文，此处当指数据的原创程度）。我们认为著作权的内涵是可变的，随着图像的创造者对数据进行可视化的角度而变化。著作权在当代的可视化信息中名目繁多，这些可

视化信息大多从属于“信息技术”（information art）这个术语，并且也深植于科学可视化之中。值得注意的是，当可视化制图的目的并不是单纯地分析数据的时候，可视化常常被视为对真相的歪曲，在学术性的可视化语境中尤其如此。虽然在这是一个简洁的视角下的合理论述，对于数据的漫像化在概念上就是不可避免的。正如费尔南达·维格斯指出的：“传统的视觉分析工具尽量减少对信息的歪曲，因为这可能会干扰冷静的分析。有没有可能，这种尽量减少视角（带来的偏差）的方式已经误入歧途？首先，创造一个真正不偏不倚的视图在总体上而言是不可能的。正如我们不改变空间距离，就无法绘制地球的表面图。”

考虑到著作权在数据可视化领域在某种程度上总是存在的，我们的分类学主要以四种方式处理数据可视化问题，这四种方式的著作权等级是按升序排列的：相片、图像、漫像以及扭曲。但是著作权并不是描述这些概念、使这些概念具有连续性的唯一方式。这种具体的方法重点在于，可视化并不是必须借助抽象的美学，而可以传递强烈的视觉隐喻，也可以在沟通语言中强调某些数据的特征。由于信息可视化是数据可视化领域的一个领域，数据可视化涵盖了科学可视化，可以说也涵盖了艺术可视化，因而这种看似不科学的分类学方法描述出了可视化领域内从更科学的一端到更艺术的一端的谱系。

数据相片

在我们的分类法中，相片是对一个数据集的最直接的复制——这是一种最接近一对一映射的数据模型。这一概念指的是李维·马诺维（Lev Manovich，2011）在既没有还原也没有直接制图的情况下进行的可视化。直接制图保留了数据的特征，而非通过雅克·贝尔坦（Jacques Bertin,1967）的位置、大小、价值、质地、颜色、方向和形式等视觉变量，将映射数据转化为一种抽象的符号。数据集中的文本在表象空间中仍旧是文本，图像仍旧是图像，语言仍旧是语言，如此等等。这样做不会侵犯著作权，因为可以从不同的角度拍摄相片，可以忠实地表现照片

的主题，同时又保留照片的每一特征，在本文所说的情况下，也就是保留了数据的特征。

数据画像

在数据画像中，作者通过反复的视觉隐喻，起到了比在照片中更加重要的作用。这里的视觉隐喻不仅仅是装饰性的视觉原素，不是爱德华·塔夫特（Edward Tufte，1983）称为“图表垃圾”的东西，而是图形的详细内容，其语义学内涵接近数据集和作者所要传递的信息。这些语义学隐喻和唐娜·考克斯（Donna Cox，2006）所说的“visaphor”含义最为接近，它们都意味着依赖于主观解释的近似数据。语义学上的视觉隐喻是在数据中以及直接映射中存在某些特征的具体证据。这些证据在图像层面得到了阐释，因而带来了不那么抽象的、更明快、更自然、更具表现力的作品。

此外，视觉隐喻具有新颖性和独特性，这些特点是深植于我们的文化之中的，然而现在我们失去了这些特征。例如，饼图是1801年威廉·费尔坡莱（William Playfair）发明的（费尔坡莱，统计摘要，对欧洲各王国的资源情况说明<伦敦，1801>）。这曾经是一种新颖的办法，随着使用的普遍化变得不再新鲜。饼图传达出一种细致的视觉隐喻，表示出一个整体的一部分。

数据漫像

我们把数据漫像当做数据画像的一种扩展形式，因为它们也利用语义学上的视觉隐喻。但是数据漫像也有自己的独有特征，如下所述。

通常认为，漫像是一个主题的形象化，通过夸大某些突出特征，来达到提高辨识度的目的。（Redman，1984）。这种所谓的辨识度是所谓的“峰位转移效应”的结果。一个习惯于表象r的个体，在看到与之相似的表象b的时候，b和r的差别越大，个体的反应就越强烈（Ramachandran和Hirstein，1999）。这样的一个参考模型r是漫像的核心部分，无论在精神上还是物质上，它总是存在的。

漫像概念在数据可视化语境中有一定的限制。正如我们解释过的，漫像基于心理上的图像给出的参照。这样的一种参照在数据可视化领域并不是理所当然的，因为即使是最直接的可视化也会给原来的非可视化数据带来一种新形式。因而，漫像在可视化语境中的应用取决于漫像所表象的参考模型的相似性，从而可以比较其不同。

如上所述，漫像与夸张和辨识度的概念是联系在一起的。在与数据有关的语境下，夸张意味着增加某个特定角度的权重，减少其他角度的权重。除了彼得·雷泰（Peter Rautek等人，2006）的著作中所指出的放大数据差异的那种方式之外，这在其他好几个方面都是可以以图形方式实现的。这会造成数据资料形式、位置、大小、颜色的视觉要素的扭曲。辨识度在数据的语境中，意味着对数据意图的说明，对来自数据的信息加以强调。虽然数据漫像和数据照片以及数据画像相比，这种形式未必是最优越的，我们仍然认为这是一种可视化的解决方案。总之，数据漫像是一种可视化的模型，这种模型用图像的方式对数据维度加以扭曲表现，用以强调同样的数据维度，或是其他数据维度。

很明显，漫像的方法之一，是通过地理位置的变形，来实现数据可视化的。这一方法自19世纪以来一直应用在统计图上。变形的统计地图是为了表现其他的数据维度。例如，世界地图上的面积统计图会根据人口或是GDP适当变形。这一方法被丹尼尔·道灵（Daniel Dorling）所简化，著名的例子就是《道灵统计图》。这种统计图不会保留地理对象的形状，而是常常用与映射成正比的圆圈的大小来表示。尽管简化了地理对象的形状，道灵已经被证明是非常有效的。这些方法的漫像性质在于强调某些数据维度（比如人口），弱化地理地图维度。这种地图是一种可供参考的模型，或者完全是被表现出来的，或者是一种精神上的建构。

数据残缺

数据漫像引入对数据的变形来表现数据，当这样的扭曲超过某个程

度，就会导致数据残缺。残缺的数据过分强调了某些特定数据，以至于传达了某些关于数据的错误信息。它们颠覆了某些视觉隐喻的语义学内涵，破坏了可视化的澄清作用，产生了难以理解的制图。然而夸张的变形给了创造更多的空间，产生了更加令人难忘的制图作品。

城市的形象可视化

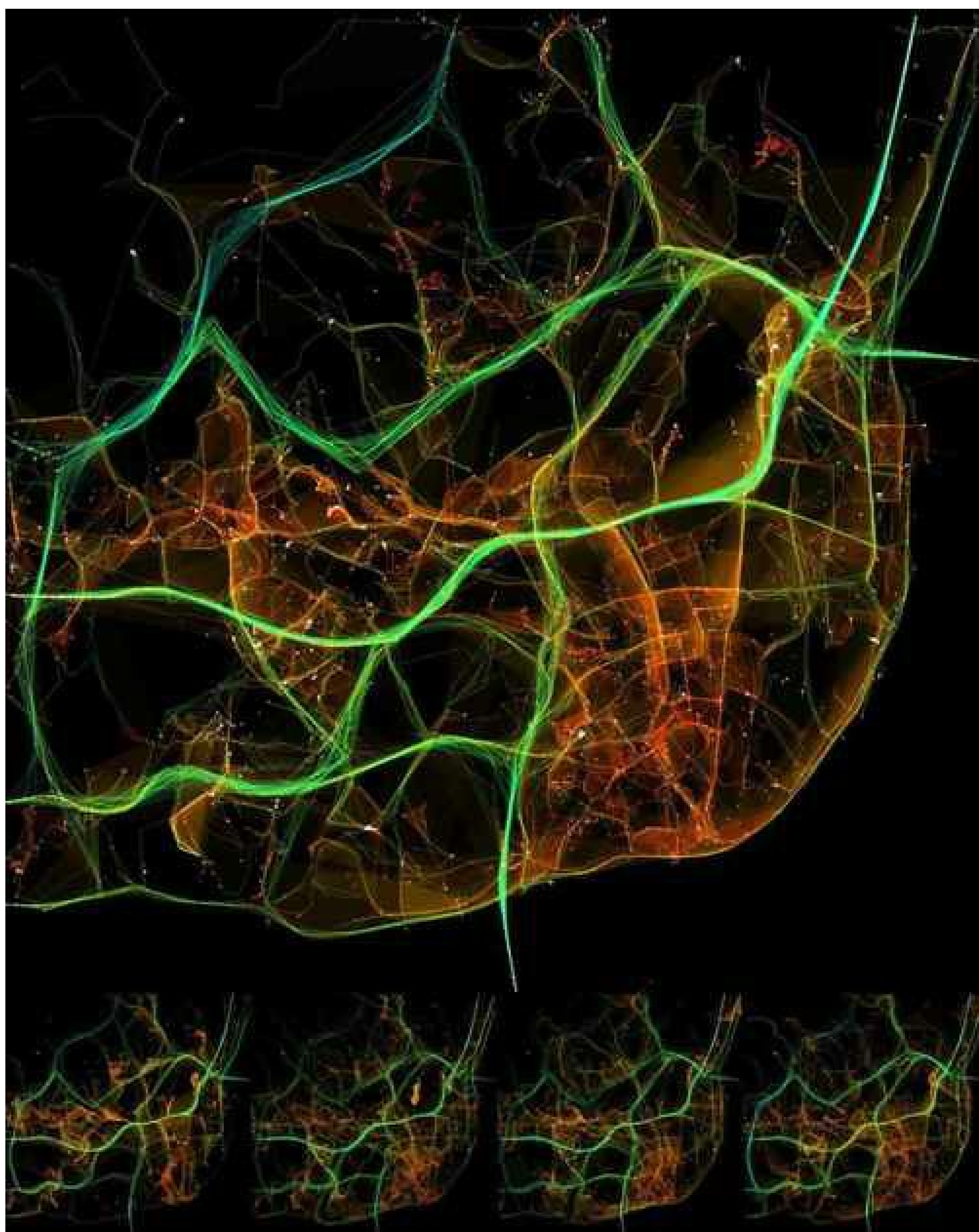
我们的案例研究以直观的、形象的方式实现了对里斯本和新加坡的城市交通系统的可视化。里斯本的数据集（数据由葡萄牙CityMotion项目／麻省理工大学提供，可视化的项目也得到了PTDC/EIA-EIA/108785/2008世界协作组织的支持，来优化移动性）包括了对城市中车辆在一个月内的GPS定位，记录了车辆位置与当前速度的信号。新加坡的数据集包括乘客的上下车信息（所谓的登入与登出），以及在一周以上的时间里在新加坡城中分别花了多少交通费。（数据由新加坡陆路交通管理局提供。可视化项目源自实时新加坡！项目，此项目由麻省理工大学的可感城市实验室以及SMART的Kristian Kloeckl负责。）

里斯本的相片

里斯本的数据集的时空精度不允许我们用可视化的方式分别描述每一天的交通模式的明确数据。因而，我们把信息压缩到单独的虚拟日之中，对每秒的数据进行分组，并通过动画加以展示。为了让交通的时间模式变得更加明显，每辆车都用一个小白点表示，在虚拟时间中会在30分钟内留下痕迹。相应的痕迹几乎是透明的，并根据车辆的速度涂上颜色。这些颜色的范围是有限的，红色和橙色表示速度放缓，绿色表示50公里／小时，青色表示更高的速度。这些轨迹往往在视觉方面不断重叠堆积，形成隆起的线条，这些线条在色块和透明的区域出现，它们（通过厚度和透明度）代表了交通的密度，（通过色彩）表示当时的平均速度。例如，狭窄的街道上的线条往往是薄薄的红色，而高速公路上的线条往往是厚厚的绿色。横跨里斯本的高速公路的颜色在交通高峰期会变成偏黄的色调。

对某个数据集的可视化通常涉及到在下面所说的系统中寻找问题：在里斯本的数据集中，那些拥挤的地区是最明显的特征。为了强调这些特征，我们添加了另外的视觉要素，将那些每30分钟就有一辆车经过的地区用很低的透明度表示。我们通过封闭回路、连接起点和终点来标明这些覆盖区域；透明度越高，颜色越接近橙色和红色，车速就越低（车速接近绿色，表明这个车速几乎是透明的。）在白天用图来表现这些区域是最大的问题，问题是要让人们更容易看到这些，而不是仅仅使用一些有色的线条来勾画这些区域——这是一个出名的难题。例如，在白天，表示市中心的线条要如何保持高亮度，城市周边地区如何在其他地区之前进入活跃状态。这不仅仅要引导观众的注意力，还要在制图中加入视觉要素，使其变得更加细致、更引入注目（见图1）。所有半透明的叠加线条与形状，使得从具体的数据要点中提取视觉要素变得不太可能，如某辆车的瞬时速度。但是这是一种直观的表现，数据的一种照片形式，并且它提供了一天之内里斯本交通进程的整体图片。

由于制图要非常细致，因而必须离线制作，之后被组合为一个动画。然而，有时制作实时的可视化图片也是有利的，让我们有机会在出现问题的时候及时发现这些问题。以下这些在新加坡的项目表现出这样的用意。



【图1】顶部：上午9点钟里斯本在高峰期的交通照片。底部：白天的城市进程：不同的区域亮度，车流用不同的颜色表示。

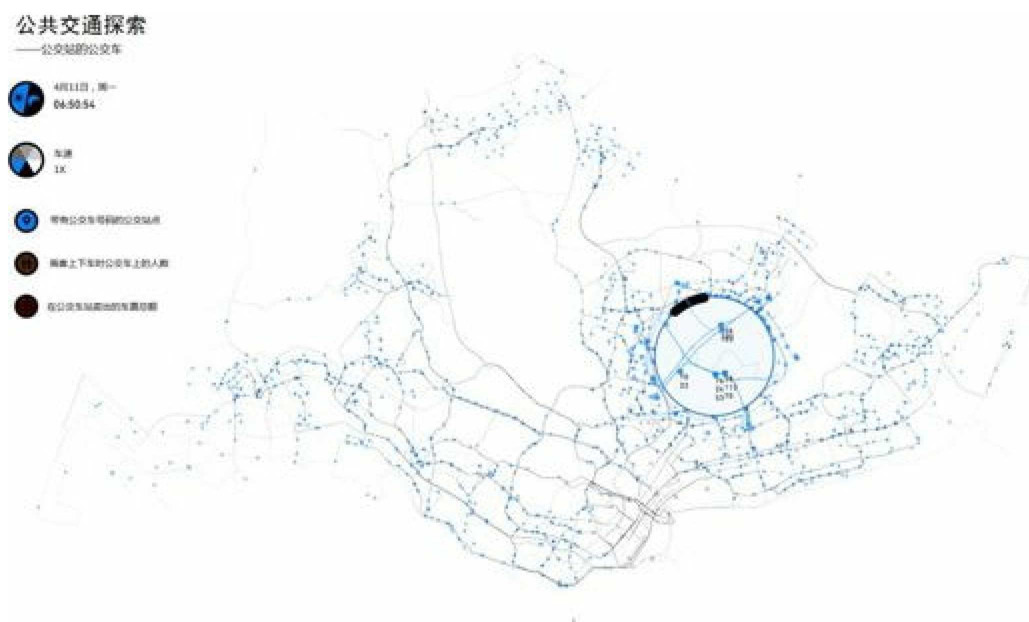
新加坡的相片

新加坡的公交网络数据集描述了公交车站的情况，但没有提供车站与车站之间的轨迹信息。尽管如此，我们认为描述这样的轨迹可以提供更加实际的制图，并且也是一个有趣的可视化挑战：表现出公交交通网络，而不仅仅表现出在车站公交车。为了达到这个目的，我们把公交车

看做是被编好程序的自动设备，仅仅对一个模拟相应输入环境（反应主体），也就是数据构成的环境做出响应。这也使得另一个目标成为可能：创建一个交互性的可视制图，在执行的时候能够运行，使得应用技术的结果能够迅速可视化。

每辆公交车，作为一个主体，只知道它下一站在哪里、到达时间以及当前的模拟时间。在执行时间内，数据被注入到模拟环境中，将每辆相应的汽车加以重新定向。每辆汽车都对下一站的情况进行缓冲，负责检索什么时候才能到下一站，什么时候能从下一站驶出。只需要知道到达时间、下一站的情况、现在的位置以及现在的时间，公交车就能开往下一站了。可以用许多方法来建构每辆车在车站之间的运动模型：第一种方法是使用非线性运动，随着这辆公交车到达下一站的时间而变化。这种方法的名字叫做“懒惰的巴士”，因为这些公交车尽管知道某个时间要停在某个车站，只要慢慢开车就可以了，但是当时间紧的时候就开得快。（两个车站之间的距离与到站时间的平方成反比。）

除了对公交车的行程进行画像般的描述之外，描述拥挤时的情况也很有趣。公交车站的密度在可视化背景上的密度比较高，这不允许在同一时间标注每个车站。在地理上，聚类算法被用来选出那些在同一时间内，有N辆车停留的最拥挤的车站。我们的算法（Ester等人，1996；Finkel和Bentley等人，1974）能够在大量数据被使用的时候在执行时间内表现良好。第一个聚类算法的结果以及“懒惰公交车”的行为模式如图2所示。小小的灰色十字代表所有的公交车站，三角形代表公交车，三角形指示着当前公交车的运行方向。只有与那一瞬间的公交车聚类相对应的车站被标记了出来。



【图2】在搜索半径44米之内聚类算法的结果，每个聚类至少有5辆公交车

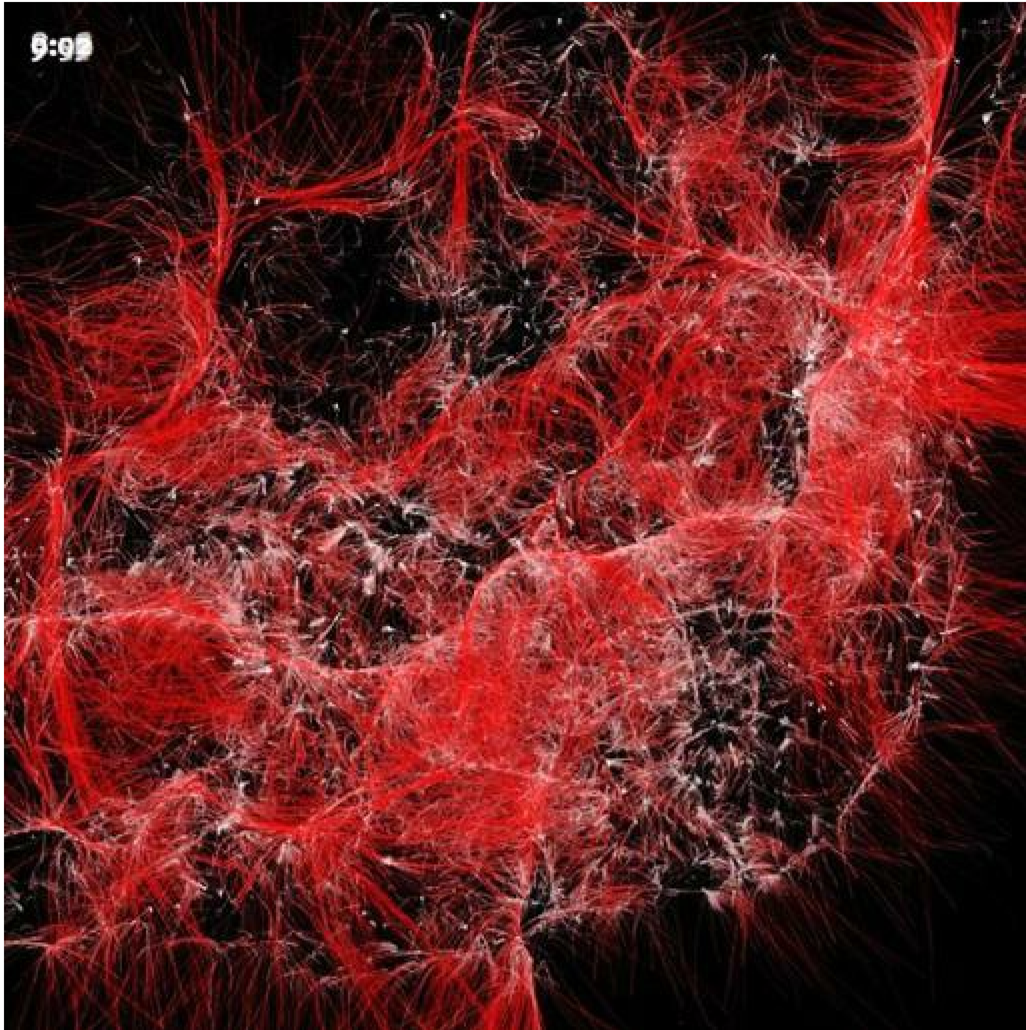
里斯本的漫像

为了绘制里斯本数据集的漫像，我们首先必须从其画像上着手，从而进行视觉上的隐喻。我们观察到的城市系统的行为是非常复杂的，但是这些行为通过形状、正常的活动周期都有各自的位置，因而任何异常的活跃都可以得到诊断。根据这种说法，并且也根据作者的审美偏好，可以绘制出里斯本的交通情况的画像，建立血管般的（交通）网络，使用循环系统来对有机体的隐喻进行探索。

计算的结构，即能够保留形式的“骨架”是有弹性的，这种结构是为里斯本的每条主要干道设计的。当道路交叉、并且在运行时相互影响其他临近区域的时候，这些骨骼就连接在一起。这使得城市能够根据交通动脉来改变其形态。弹性装置能够随着某些在受到作用力的时候产生的反作用力而改变形状。如果这些数据的变化就是作用于那些弹簧上的力，那么我们就可以实现在城市形态之间的平滑转换，当这些作用力平息的时候，也能够自然地回到初始形态。这些形态的变化通过与普通的地理映射进行对比（也就是我们所说的漫像，即参考模型），转化了城市中实际的感知距离。这种类型的映射是一个典型的可视化漫像，因为

它是一种有关距离的漫像——这种漫像以距离统计图的形式出现是比较常见的，从一个精确定义的原点在旅途时间的函数中对地理距离进行一对一的映射。在我们所说的情况下，是想要呈现旅行时间如何与城市主动脉相关联，同时如何改变这些距离，如何改变城市的整体形态。

我们正在处理一个相当复杂的弹性的物理系统。如前所述，每个弹性装置都能够根据数据扩大或者缩小，在我们所处理的情况中，每个弹性装置又与其他的弹性装置相联系，形成一个骨架。因而，每个弹性装置不能自由改变其形态，它还同时被其他弹性装置和数据所影响。经常有许多相反的作用力作用于相同的弹性装置，这里存在着很高的风险——当骨架无法辨认出城市图像的时候，就会崩溃，进而产生有缺陷的图像。因而，我们必须确保其暂时的稳定性，来减弱那些小变化，给系统时间来适应新变化，来避免崩溃的发生。这样，数据被均分在一天当中，每隔10分钟仅仅出现一个小时的聚合，从而来对系统产生影响，制造可视化效果（如，从3:20到4:20，再从3:30到4:30，等等）。系统在以下几个情况下会被激活：车流量变得更大，使得（线条）增厚，车速变高，使得（线条）变长（反之亦然）。我们选择后面的行为来传达整个城市感知距离的总体印象。当交通的速度变高的时候，（系统的）这种行为就会减少，当在高峰期城市面临拥堵问题的时候，这种行为就会增加。“血管”也被标记为相应的颜色：速度越低，颜色越暗，表示循环变慢、血流停滞（参见图3）。车辆的可视化会带来天然的美学制图，它是天生的视觉隐喻，每当高峰期交通拥挤压力大的时候，就会产生脉动。动脉的压缩和扩张以及整个城市的漫像显示了感知距离以及超出平均速度的交通情况。



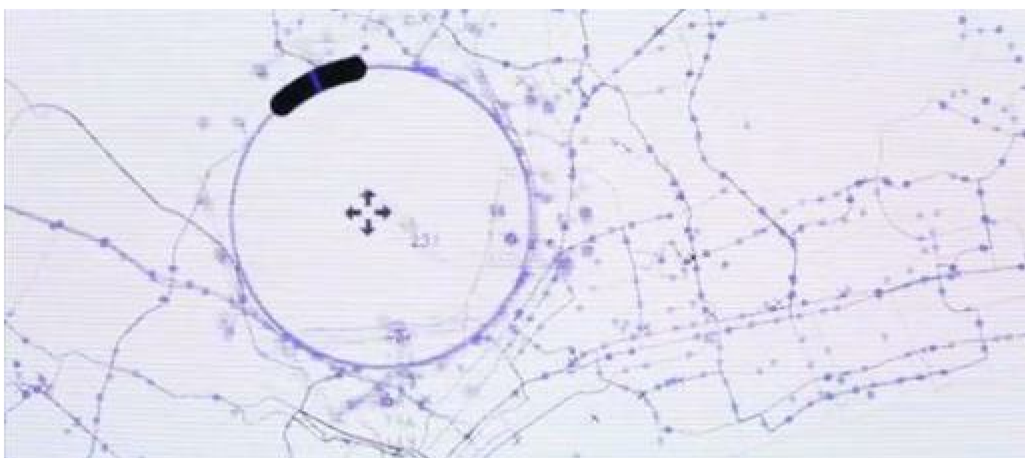
【图3】

在早晨7:04（左）和早晨8:44（右）里斯本的“血管”。这恰好是早高峰之前和之后的两个时间点。可以观察到，早高峰开始的时候，里斯本的主干道要负载大量的车流，但是没有产生交通拥堵的问题，也就是说“血管”是收缩的。上午8:44，血流速度降低，这使得“血管”和整个城市呈扩张状态。

另一个典型的体现出漫像风格的技术，是乔治·弗纳斯发明的鱼眼视角（1986），此后这一技术被应用于绘制地图（Keahey和Robertson，1996）。这项技术放大我们关注的地理区域，而对外围地区加以扭曲，以便在背景上保留其他区域。对于统计图来说，主要的区别在于统计图

的工作方式不同：统计图在一个维度上进行扭曲（地理维度），是为了强调另一个维度（比如GDP）。而在使用鱼眼视角的情况下，同一个维度（地理维度）上出现了扭曲，为的是强调同一个维度内的某些要素（关注的区域）。此外，鱼眼视角在使用设备绘制数据画像的时候也是一个视觉隐喻：它隐喻了放大镜。

新加坡的数据集漫像采用了这个方法。我们使用了一个数据镜头，从而在不脱离整体语境的情况下，解决展示在街道层面展示每个公交车站的问题。然而，经典鱼眼镜头的放大倍数无法满足这种可视化的要求。为了实现这一点，我们发明了一种新型的镜头，能够改变放大镜的运行方式（通常的做法是在飞机上向半球进行投影来绘制地图），将之转变为从飞机上向一个非常扁的椭圆球体进行投影（参见图4）。



【图4】

我们设计的数据镜头是一个交互式的可视化工具，让用户能够发现不同层次的信息，这些信息可以揭示新加坡公交网络的更多细节。镜头可以在城市交通网拖拽，也可以聚焦在某一组公交站台。用户可以修改镜头的属性（位置、大小和缩放级别）来区分不同的公交车站，哪怕车站之间距离很近也可以区分。此外，镜头还让我们能够在数据集中通过过滤某些类型的信息层次，来发现某些特性：在每个车站分别有哪几路公交车、每辆车上有多少人、在任意一站每位乘客花费的总费用。用户

可以在这些信息层次之间切换，来探索公交线路、车站、等车时间、载客量和车费之间的联系。

这一装置可以带来更多有趣的浏览空间，用户可以把界面引导到一簇点中，顺畅地浏览并理解其每个组成部分。

评论

数据漫像旨在使用艺术可视化的做法，通过平衡科学的客观性与对科学可视化进行描述，推进当代信息可视化的边界。这种平衡通过以明确的意图设计的方法，绘制有关数据的漫像得以维持——它能够有效地与观众沟通。

数据漫像用隐喻的方法，此外还包括照片、画像、漫像、有缺陷的可视化等方式得以形成。照片主要关于直接的映射或者可视化，并不包括任何缩减。画像和照片不同，它引入了强烈的视觉隐喻。视觉隐喻有语义学上的意图，构成了数据的某些特征以及直接映射的比喻性的证据。这样的证据在地理层面得到了阐释，带来了不那么抽象、因而更具表现力的制图。数据漫像拥抱了视觉隐喻的概念，用夸张来对其加以延伸，以强调数据的某些方面。这样的夸张可以通过引入粗糙的变形——也就是有缺陷的数据。

我们把漫像的概念应用于城市的可视化。我们相信，通过使用一种漫像式描绘城市的方法，通过传递更加具体的信息，可以更贴近观众。例如，城市居民对此已经有了某些心理印象——如果我们可以把这个印象和理想模型的表现加以比较，我们就有了一个扭曲的表现形式，这个时候就可以充分利用漫像。此外，这些扭曲也可以是可视化模型的自然结果，随着复杂数学问题的近似策略而产生。比如在我们的工作中，代理或者是物理模型，能够顺利地适应数据，这些模型比那些迅速做出反应的模型有更广的适用性，使得我们能够把它们再用于其他城市，甚至其他类型的数据。事实上，这些模型对于漫像的方法来说是理想状态下的，因为根据格雷格·尤德尔曼（Greg Judelman, 2004）的说法，它们更

接近自然美学，在解码复杂情况的可视化语境下得到更经常的运用。比如在自然界中，复杂性是广泛存在的，在自然的驱动下的系统的应用对于这些城市来说似乎是一种自然的漫像，描述了不同尺寸的信息，同时又保留了其核心内容。由于这些，我们相信漫像能够在信息可视化中起到重要作用，增加效率、进行难忘的交流、表现城市，并且拉近城与人的距离。

流动性的计算模型：来自移动电话数据的视角

作者　菲利普·赫费尔（Philipp　Hövel）　菲利波·西米尼（Filippo Simini）　宋朝鸣　艾伯特－拉斯洛·鲍劳巴希（Albert-László Barabási）

译者　曹海元

近年来，由于网络科学的崛起，产生了越来越多的数据。（Albert和Barabási，2002；Lazer等人，2009）。可利用的网络涵盖了众多领域和机制，包括了蛋白质间相互作用、食物网络、信息流、交通运输、社会交互等其他更多方面。特别地，随着数字时代中可获取数据集数目的增长，可以在更为详细的层面上进行社会网络研究。其中包括了反映友谊、消费行为和流动性的网络。

在社会环境下，对通常称之为“大数据”的分析可以用来推测人类行为（Barabási，2005；Eagle和Pentland，2006；Onnela等人2007a；Eagle和Pentland，2009）。进一步说，经验数据是我们能够理解、预测和建模诸如人口流动这样的日常生活各个方面的关键。从移动电话使用中获得的这一类数据集特别适合作为研究流动性的切入点（González和Barabási，2007；González等人，2008；Wang等人，2011；Bagrow等人，2011；Bagrow和Lin，2012）。例如本章中使用的经验数据集包含了来自某一电话公司一千万用户的匿名呼叫数据记录（CDR）。经验数据是我们能够理解、预测和建模诸如人口流动这样的日常生活各个方面的关键。用户每打一个电话或者发送一条短信就会生成一条匿名CDR。它包含了匿名呼叫方和接听方的标示符，打电话或者发短信的日期时间以及通信者的位置信息。位置信息是依据处理这次通话的最近通信基站位置得到的近似值。类似的数据集已经被用来研究通信网络中的地理信息以及与空间有关的通信（Lambiotte等人，2008；Blondel等人2010；Expert等人，2011）。可以举出一个例子，在比利时境内根据使用语言

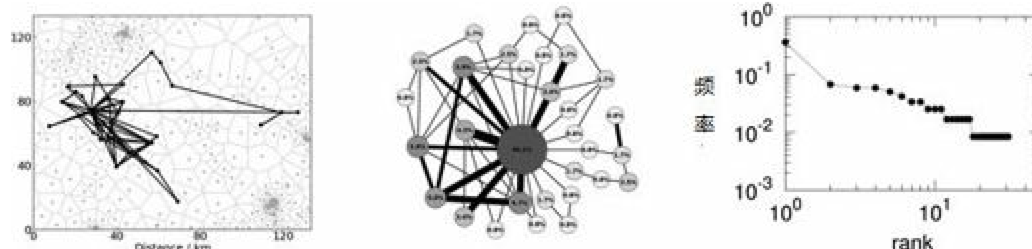
可以很明显地划分为荷语区和法语区。这个划分是依据移动电话通信网络中进行的社群探测，即通过把网络节点（用户）整理成有较多内部链接、较少外部与社群间链接的集群的方法。

本章的目的是基于对移动电话数据的分析给人口流动提供一个视角。我们提出了以下问题：在日常生活方式中我们的如何预测我们在日常生活中的行为（Song等人，2010b）？人们出行所表现出数学定律和函数关系，也被称为标度定律，是怎样的（Song等人，2010a）？用于描述观察的行为模式的模型是由什么所组成（Simini等人，2012）？

为了回答这些问题，我们不仅考虑了数据在流动性方面就个人移动电话用户的大致位置（电话从哪里拨打）所提供的信息，也考虑了数据集的社会信息（谁给谁打了电话）。进一步而言，我们的分析在距离上跨越了几个量级，从城市到地区，再延伸至全国。

人口流动的可预测性

为了了解在本研究中所用的数据，图（1）左显示了单个用户在一个月观察周期内的样本轨迹。通过记录每次用户打电话的位置（黑点），可以看出这位特定用户拜访了N＝31处不同的地方。基站的位置用灰点表示，灰线标注了通过Voronoi网格（Voronoi lattice）推导得出的近似信号接收区域。通过最靠近不同基站的位置之间的边界确定了每个网格的范围。换句话说，所有处在某个基站的Voronoi单元格内的位置都拥有一个共同点，它们与这座基站之间的距离比其余基站都要短。这也是移动网络供应商用来确定基站接收区域的标准方法。黑色线条代表了通过时间顺序连接用户通话位置形成的轨迹。



【图1】左图：某位匿名手机用户在一个月观察期内访问了31个不同基站附近的地区所产生

的轨迹，期间用户拨打了总共L＝119个电话。每当用户拨打一个电话时，最近负责转接的基站被记录了下来，确定了用户的大致位置（黑点）。灰线表示了Voronoi网格，大致标明了每座基站的接收区域（灰点）。黑线对应被记录的用户在基站之间的运动。

中图：同一个用户的移动网络。节点的尺寸对应于用户在每个基站临近区域中通信的比例，而连线的宽度和不同基站之间直接移动的频率成正比。

右图：位置的排名vs访问的频率（Zipf图）。

在31个地点之中，有大量的地点只被访问了几次，但只有一小部分地点是用户会经常返回。这个典型的行为被抽象地描绘在中间图展示的移动网络中。在那里，每个网络节点对应着一个位置，如果在两个位置接连通信了两次，则这两个节点被连在一起。节点的尺寸反映了源自每个基站的通信事件比例。在这个可视化中人们可以轻易地确定出一个连接大量访问数较少节点的中央枢纽。依照访问的频率将不同地点进行排序，它们服从一个幂律的分布，如同图（1）右图所绘。这被称作Zipf定律（Zipf，1946）。

用回转半径可以进一步量化流动性。这个量通常是由所有记录的位置相对轨迹质心的平均距离计算得到的，它提供了出行的特征长度。

数学上，对在L个通信事件中位置被记录为$\vec{r}_1,\ldots,\vec{r}_L$的每位用户而言，他们的回转半径$r_g$被定义为：

$$r_g = \sqrt{\frac{1}{L}\sum_{i=1}^{L}\left(\vec{r}_i - \vec{r}_{cm}\right)^2} \qquad (1)$$

【图2】

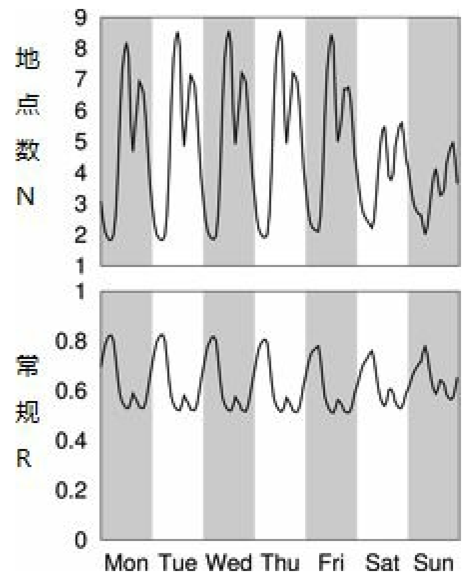
其中质心被定义为：

$$\vec{r}_{cm} = L^{-1}\sum_{i=1}^{L}\vec{r}_i$$

【图3】归一化的回转半径rg的分布。数据：2009年2月

图（3）所示为对所有用户记录进行计算得到的回转半径rg的分布。

曲线展示出了长尾分布的形状。这预示了大多数的出行是短途的，只覆盖了短距离的范围。同样，只有部分用户展示了数十或数百公里特有的出行距离。



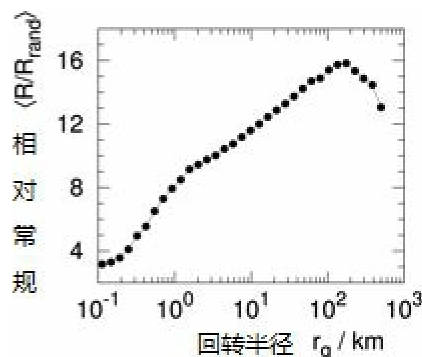
【图4】流动性的规律：一周的每天，以小时为时间间隔，R表示在某用户最常访问的地方找到其本人的比例，N表示不同地点的数量。数据：2009年5月，对一组每天进行12到100件通信事件的用户数据进行按日平均计算。

除了出行特征长度的分析，我们还利用了嵌入在数据组中的时间信息（即通信事件发生的时间）。这解决了预测性的问题。简而言之，我们累积了在几周时间内，每周168个小时中每位用户的所有通话记录。我们提取出了每位用户的一个地点列表，即他／她在周一凌晨12:00到凌晨1:00，凌晨1:00到凌晨2:00，直到周日深夜11:00。至半夜的所有时间区间内和某人通信的地点信息。接着我们为168个列表中的每个确定了主要地点，即处理最多通信任务基站的地点。这个基站记录的任务相对所有通信任务的比例记为：

$$R1=\frac{\text{出现在主要地点的次数}}{\text{出现的总次数}} \quad (2)$$

这个公式提供了在一周中每个小时，i＝1,…,168衡量规律性的工具。举个例子：如果某位用户在特定的某个小时内在一个地点出现8次而在另一个地点出现两次，那么我们得到了一个衡量规律性的值Ri＝8/

（8＋2）＝0.8。图（4）描述了在一星期168小时中访问的平均地点数目N（上图）和平均规律度R（下图）。这里，平均值来自于一组每天进行12到100件通信事件的用户数据。这个曲线肯定了人们直觉上可能的期望：在早晨的时段人们去的地方最少且最规律，而在从家到工作地点的通勤时间段去得地方最多且最不规律。这个效应在每个工作日重复出现，但是在周末最不显著。



【图5】平均相对规律度vs回转半径喻示了具有较大回转半径的用户具有相对较高的规律度。

一个和开头介绍的回转半径与规律度有关的问题是：具有较长特征旅行距离的用户（Rg值较大）在他们的行为中展示了较高还是较低的规律度？

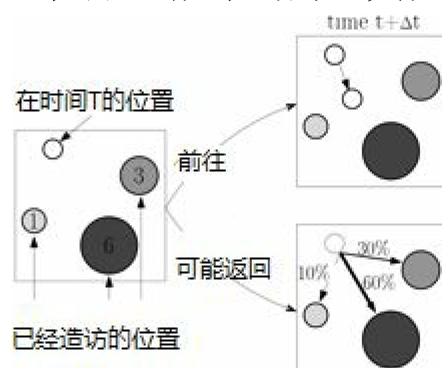
为了解决地点总数上的差别，我们考虑了相对规律度R/R rand，其中R rand＝1/N作为相对标度。这个情况中假设了用户以随机但均匀的方式访问了N个地点。有人可能会期望那些不远游的人会更有规律，但是评估数据之后发现了相反的趋势。图（5）表明了每周平均的规律度相对回转半径变化的结果。与直觉相反的是，我们发现回转半径较大的用户拥有相对较大的规律度。换句话说，出行距离较长的用户会变得更有规律。

借助信息论和统计物理的技术，我们可以估算出每位用户在某地出现的最大可预见性。涉及到熵和法诺不等式（Fano's inequality）的细节请参见Song等人（2010b）以及其中的参考文献。它的主要结果有两

条：（1）最大可预见性的在0.93处呈现出狭窄峰值，意味着随机挑选目的地的次数不超过7％。（2）令人惊讶的是，可预见性并没有在那些有着较大回转半径的用户身上呈现出急剧降低的趋势，但是它在0.93处饱和。这说明无论一组人去的地方有多远，在93％的次数里他们的目的地是可预知的。

人口流动的数学描述

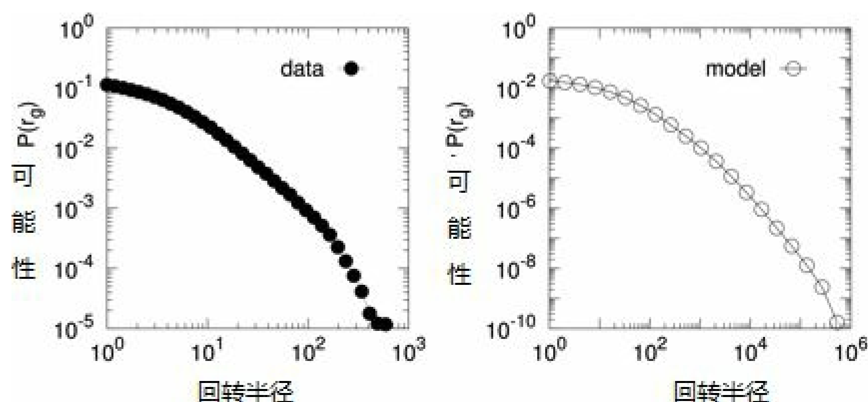
到目前为止，我们讨论了人们出行的规律性与特征长度方面的经验性证据。在接下来，我们会展现能够重现经验发现的简单模型。首先，我们关注到个人移动的随机模型以及考虑到回转半径后它的精确性。接着我们引入了一个从第一性原理推导出的，关注迁徙模式的模型。



【图6】个人移动模型的示意图：左盒中显示的是在t时刻的出行路程构型n，表示了用户在当前及曾经去访问过的地点（这里：N＝4，包括了当前的地点）。画在每个地点圆的大小正比于访问该地点的频率。在t＋Δt时刻（其中Δt选自一个长尾分布）某个概率下用户可以从他／她当前位置去往一个新的地点，其移动的距离基于一个长尾分布（探索；右上），除此之外用户会回到之前访问过的地点，去的地方是按照他／她之前访问地点的频率来进行选择的（偏好返回；右下图）。

个人移动模型建立在对我们日常出行的经验观察，其观察结果主要是很多近距离、短等候时间而且频繁去的地点，而大量长途旅行与相邻旅行之间的长时间等候是很少出现的（González等人，2008）。数学上，例如等候时间Δt的概率分布P（Δt）可以写成长尾分布P（Δt）～|Δt|-1－β其中0≤β≤1。这个公式反映了这样的数学意义：两次出行之间长

时间等待次数要少于短暂停留。移动距离分布描述了由访问基站之间距离给出旅程长度的统计结果与α次幂有着相似的形式。基于这两项发现可以简单构建出一个连续时间随机行走模型（Brockmann等人，2006），但正如Song等人（2010a）所示，这并不能描述人口流动的所有方面。

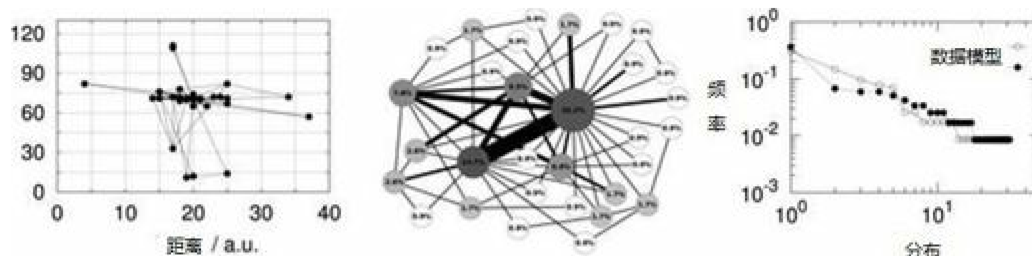


【图7】从移动电话数据推算出回转半径Rg的归一分布（左图，观察周期：1000小时），个人移动模型（右图，任意空间单位）。该图重制自Song等人（2010）的图4（a）与（c）。

除了所选的等候时间与移动距离满足的标度定律，还应该包括在图（6）中示意性描述的两个普适机制。第一个机制是探索。在一段等候时间Δt之后，个体到达了一个前所未知的地点。等候时间和移动距离都按照经验上证实的长尾分布选取。事件的发生存有一定的概率，其中N还是代表了之前访问过的地点数目。也就是说，当某人已经去过很多地点后，探索新地点的可能性会变得更小。参数ϱ∈（0,1］与γ≥0可以由数据决定。相应的，用户有着$1-P_{exploration}$的概率回到之前已经访问过的地点。因为它是依赖于概率$P_{exploration}$的结果，所以当某人已经访问过许多地方后，更倾向于回到之前的地方。此外，从已经走访的地点中选择要返回的地点，其概率正比于走访的频率，也就是说，走访频率越高的地方，越有可能被选作返回的地方。这种现象叫做“偏好返回”。

考虑本章之前讨论过的回转半径（见等式1）可以证实这个模型。结果见图（7）。左图与右图分布显示了实测数据与模型预言的结果。曲

线的形状和标度指数很好地支持了所提出模型的假设。



【图8】左图：根据个人移动模型预言的运动轨迹（α＝β＝0.6,γ＝0.2,ϱ＝0.4）。模型中的用户行为与图（1）中描绘的类似，在L＝116个通信事件期间探访了N＝31个地点。

中图：同一个用户的移动网络（与图（1）的布局相同）

右图：地点的排名vs探访的频率（Zipf图），图（1）中的经验数据添加为实心圆点。

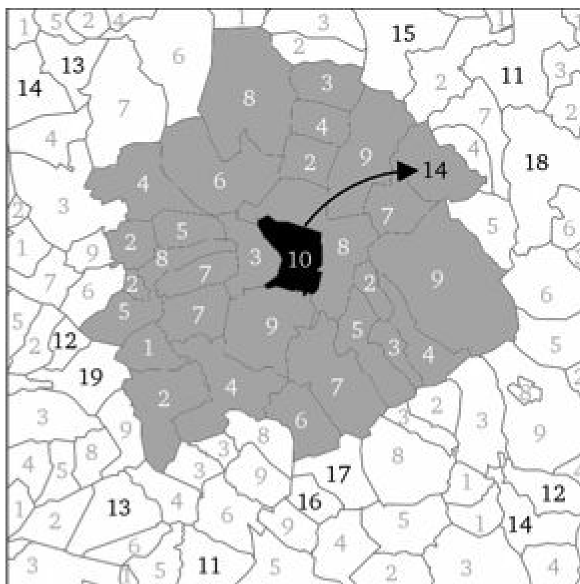
为了证明个人移动模型很好的与手边的数据相符合，图（8）显示了模型预测的单个用户案例（α＝β＝0.6,γ＝0.2,ϱ＝0.4）。与图（1）的比较已经提示了模型结果与移动网络和轨迹的观察结果有非常相近的移动特征。模型同样定量地符合Zipf图（图右图），其中展示了经验数据和模型预测的结果。

除了以上展示的结果，个体移动模型也能重现出其他特性，例如对在特定数量的已访问地点找到用户的概率，或者随时间增长缓慢的回转半径。这里超慢增长指的是相比于连续时间随机行走模型中标准参考案例的期望增长率要慢很多（Brockmann等人2006；详情见Song等人2010a）。

普适迁移模型

个人移动模型几乎都带有固有的波动，很好地再现了人类移动的标度性质，还能够制定一个无参数的模型，用来描述流动和迁移模式。它只基于人口密度。因此，它不需要任何先验的信息，诸如等候时间或跳跃幅度分布。这个模型被称为“辐射模型”，因为它是由以下的物理概念推导得出的（Simini等人，2012）：粒子（出行者）从一个源头（起始位置）向外四散，可以被环境以一定概率吸收（在附近地点终结行进路线）。粒子行进距离（通勤路程长度）依赖于材料的厚度（附近地点的

人口密度）。



【图9】辐射模型的示意图：城市内的数字代表着相应的益处，意思是区域内企业给予的最高报价。从中央区域（黑色）出发的个体到附近具有更高益处的区域（箭头）。

因此，辐射模型的目标是估算通勤的流量，例如单位时间内穿梭在国家中任意两个区域间的平均旅行人数。模型从地点i和j各自的人口数m_i和n_j开始。进一步，通过计算在以地点i为圆心，半径为r_{ij}的圆内总人口S_{ij}来考虑地点i和j之间的累计人口数目。模型预计的从地点i到j间的平均流量如下式所示：

$$\langle T_{ij}\rangle = T_i \frac{m_i n_j}{\left(m_i+s_{ij}\right)\left(m_i+n_j+s_{ij}\right)}, \tag{3}$$

【图10】

其中$Ti \equiv \sum_{i\neq j} Tij$表示从地点i出发的总人数。考虑到总共有$N_c$位通勤者，那么总人数可以简单表示为$T_i = m_i N_c / \sum_i m_i$其中$\sum_i m_i$是研究中整个区域的人口数目。

图（12）提供了关于辐射模型的一个示意图。描述如下：位于某个

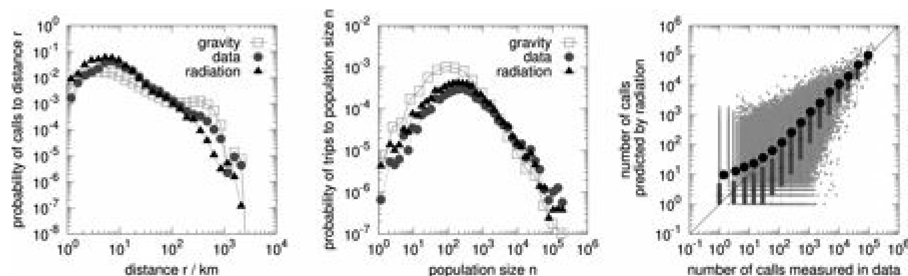
特定地点的个人会移动到附近相比他／她原先地址提供更多益处的地点，而每个地点的益处（图12中显示的数字）表示了当地企业提供的最高待遇，这个数字与每个地区的人口数量成正比。在图中，示例的情况为个人从市中心（益处为10，黑点）被预测会流通到益处为14的城市区域，如箭头所示，因为它是拥有较高益处的最近地点。灰色区域标示的是在模型（3）中出现的半径对应的圆形。其内部所有的区域提供的益处都比原始地点要少。

Simini等人（2012）展示了辐射模型在预测迁徙模式上要优于其他模型，诸如广泛使用的引力模型。它决定了从地点i到j旅行的个体数目T_{ij}服从如下的分布：

$$T_{ij}=\frac{m_i^{\alpha}n_j^{\beta}}{f\left(r_{ij}\right)}, \qquad (4)$$

【图11】

其中的阻尼函数通常被假设为幂函数或者指数函数，f（r_{ij}）中的α和β是可调指数。其名字来源于与牛顿引力定律的类比。假设两座城市间的迁移类似于引力系统，引力相互作用中物体的质量被替换为不同地点的人口密度。相应地，在距离更近和人口更多的地方之间具有更强的相互作用，即人口流动量。然而，这个模型有着几个局限性，例如一个严格的推导过程，对阻尼函数和参数的恰当选择，对数据拟合参数的要求，在庞大人口规模极限下的无界性，以及对区域i和j之间地形差异固有的忽略。辐射模型可以克服所有这些局限性。



【图12】左图：通信距离的概率分布。中图：人口数为n的地区通信概率分布。右图：比较

不同城市地点间测量与模拟人流的结果数据：在4周内不同城市区域用户通信数目，总计4,336,217位用户通信了38,469,153次。通过数据积累获得了不同城市对之间的电话总数。图3（g）至（i）复制于Simini等人（2012）。

一般来说，扩散模型可以用于不同情况中，诸如长期迁移或者运输驱动过程，譬如货物运输。这里我们专注于移动电话数据赋予它的预测能力。图（12）介绍了关于数据结果、引力模型和辐射模型对比得到的结果。左图描绘了两个距离为r的城市之间发生一次通信的概率分布。当引力模型在小距离与大距离条件下和实际数据出现分歧时，辐射模型与实际数据符合地更好。同样的情况也发生在旅行目的地是人口数目为n的地点的概率分布上。最后，右图显示了所有城市地点对之间流量的数据值与它们预测值之间额比较（灰点）。黑点对应的是每个区间内预测通信数目的平均值。蜡烛图柱形表示了9％和91％的百分位数值。每个区间与对角线的内部重叠整合在一起变为一条细黑线。

总结

我们论证了在我们的移动电话数据案例中，大数据是如何被用来分析人口在几个不同量级上的移动行为。我们确定了日常生活中移动满足的标度定律和规律性，提出了简单模型可以在很大程度上再现这些经验的发现。所讨论标度定律的普适性和模型的实用性表明了，基础设施规模对人类行为的相关性和重要性。

致谢

Phillip Hoevel感谢德国学术交换计划（DAAD）通过博士后基金提供的支持。

参考文献

1．Albert, Réka, and Albert-László Barabási. 2002. "Statistical Mechanics of Complex Networks," Reviews of Modern Physics 74, no. 1 (January): 47-97.

2．Bagrow, James P., and Yu-Ru Lin. 2012. "Mesoscopic Structure and

Social Aspects of Human Mobility," PLoS ONE 7, no. 5: e37676.

3．Bagrow, James P., Dashun Wang, and Albert-László Barabási. 2011. "Collective Response of Human Populations to Large-Scale Emergencies," PLoS ONE 6, no. 3: e17680.

4．Barabási, Albert-László. 2005. "The Origin of Bursts and Heavy Tails in Human Dynamics," Nature 435, no. 7039: 207-211.

5．Blondel, Vincent, Gautier Krings, and Isabelle Thomas. 2010. "Regions and Borders of Mobile Telefony in Belgium and in the Brussels Metropolitan Zone," Brussels Studies no. 42 (October 4).

6．Brockmann, Dirk, Lars Hufnagel, and Theo Geisel. 2006. "The Scaling Laws of Human Travel," Nature 439, no. 7075: 462-465.

7．Eagle, Nathan, and Alex Pentland. 2006. "Reality Mining: Sensing Complex Social Systems," Personal and Ubiquitous Computing 10, no. 4 (May): 255-268.

8．Eagle, Nathan, and Alex Pentland. 2009. "Eigenbehaviors: Identifying Structure in Routine," Behavioral Ecology and Sociobiology 63 (May): 1057-1066.

9．Expert, Paul, Tim S. Evans, Vincent Blondel, and Renaud Lambiotte. 2011. "Uncovering Space-Independent Communities in Spatial Networks," Proceedings of the National Academy of Sciences 108, no. 19: 7663-7668.

10．González, Marta C., and Albert-László Barabási. 2007. "From Data to Models," Nature Physics 3, no. 4: 224-225.

11．González, Marta C., César A. Hidalgo, and Albert-László Barabási. 2008. "Understanding Individual Human Mobility Patterns," Nature 453, no. 7196: 779-782.

12．Lambiotte, Renaud, Vincent Blondel, Cristobald de Kerchove, Etienne Huens, Christophe Prieur, Zbigniew Smoreda, and Paul Van Dooren.

2008. "Geographical Dispersal of Mobile Communication Networks," Physica A: Statistical Mechanics and Its Applications 387, no. 21: 5317-5325.

13. Lazer, David M. J., Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. 2009. "Computational Social Science," Science 323, no. 5915: 721-723.

14. Onnela, Jukka-Pekka, Jari Saramäki, Jörkki Hyvönen, Gábor Szabó, Marcio Argollo de Menezes, Kimmo Kaski, Albert-László Barabási, and János Kertész. 2007a. "Analysis of a Large-Scale Weighted Network of One-to-One Human Communication," New Journal of Physics 9, no. 6 (June): article 179.

15. Onnela, Jukka-Pekka, Jari Saramäki, Jörkki Hyvönen, Gábor Szabó, David M. J. Lazer, Kimmo Kaski, János Kertész, and Albert-László Barabási. 2007b. "Structure and Tie Strengths in Mobile Communication Networks," Proceedings of the National Academy of Sciences 104, no. 18: 7332-7336.

16. Simini, Filippo, Marta C. Gonzalez, Amos Maritan, and Albert-László Barabási. 2012. "A Universal Model for Mobility and Migration Patterns," Nature 484, no. 7392: 96-100.

17. Song, Chaoming, Tai Koren, Pu Wang, and Albert-László Barabási. 2010a. "Modelling the Scaling Properties of Human Mobility," Nature Physics 6, no. 10: 818-823.

18. Song, Chaoming, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. 2010b. "Limits of Predictability in Human Mobility," Science 327, no. 5968: 1018-1021.

19. Wang, Dashun, Dino Pedreschi, Chaoming Song, Fosca Giannotti,

and Albert-László Barabási. 2011. "Human Mobility, Social Ties, and Link Prediction," in Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 21-24, San Diego, CA. New York: Association of Computing Machinery, 1100-1108.

20. Zipf, George K. 1946. "The P1 P2/D Hypothesis: On the Intercity Movement of Persons," American Sociological Review 11, no. 6 (December): 677-686.

数据中的城市，城市中的数据

作者　凯尔·格雷科（Kael Greco）

译者　常莹

2007年年初，一群Google Earth的用户在圣地亚哥发现了一件奇怪的事情。在使用最新发布的卫星影像进行俯瞰时，这些“圈椅里的全球定位系统”（译注：意即次系统令用户足不出户便可进行全球定位）注意到有一组建筑在从上面看下来的时候会呈现出一个令人费解的形状——纳粹的十字记号。这个发现像病毒一样扩散开来——早在病毒传播这个概念出现之前——主要的新闻媒体都报道了此事。人们很快发现，这组综合设施实际上是由美国海军在1967年建造的，巧合的是它周围的公路也是以二战相关的地点命名的。当时所有的人都可以通过一个网络链接去查看它，这个基地的俯视图引发了公众的强烈抗议，最终在政治压力之下，海军耗费了60万美元进行了一个改建项目来改变这组建筑那令人厌恶的外形（Perry）。一位发言人说：“我们不想和那个代表纳粹的、引人憎恶的十字符号有什么关系。”海军声称这组建筑的实际形状和排列完全是无心之举，是监管疏忽的后果，他们为此感到羞耻。但无论此举是否出于有意，有一点都很清晰，那就是这项工程的规划者、设计师和建造者都完全没有预想过会有人以上帝一样的视角去俯瞰完工后的建筑。这件趣事听来不但可笑，同时也揭示出一个特别重要的概念：新的视角可能会彻底再造我们对一个地方的了解以及我们与它的关系。我们确实已经可以以全新的视角来观察城市，在这股风潮之下，Google已经重新加工和整合了城市的数据——这个例子，就帮助人们发现了这个星球上由政府资助建立的最大的纳粹十字星。

当然，并非所有的城市数据集都像卫星影像一样在本质上就是可见的。有很多数据需要配合城市空间信息，才能够理解和应用——也就是说，它们要在城市的背景之下才能发挥出作用和力量。这种力量的一个

实际例子——感谢爱德华·塔夫特（Edward Tufte）在传播数据知识方面的努力工作（Tufte，24）——就是约翰·斯诺（John Snow）在1854年绘制的霍乱爆发地图。19世纪中叶的时候，随着人口集中区域的人口密度不断增长，城市不可避免地面临着公共健康问题，规划者和理论家们希望能够解决这种不断增长的健康威胁。在1850年代中期，最受人关注的就是霍乱的传播。医学界尚未接受微生物学的基础理论，但是通过绘制苏活区因霍乱死亡病例的地图，斯诺可以清晰地向人说明，霍乱传播的媒介不是被污染的空气，而是被污染的水和食物。

数据过载与维度的灾难

就像科技的进步一样，可视化与映射技术的应用也不可避免地与极大的数据量关联在一起，就像詹姆斯·科纳所说的："……映射技术为我们提供了各种可能性；它一次又一次地重塑了领土的形态，每次重塑都带来各种不同的、全新的结果"（Coner 2013）。这种观念与不断传播的"大数据"的案例一起，催生出一种令人兴奋的信念：一切事物都可以通过计算被描述、操纵、模式化以及模拟。在大数据的时代，总量为王，任何维度上的补充都会在代表性方面带我们向完美更进一步。博格斯（Jorge Luis Borges）的《论科学的精确性》是一篇讽喻小说，描绘了一个追求一比一复制的幻想世界，这个故事已经快被我们遗忘了。我们不再对全面地刻画与模拟这个世界的前景感到畏惧——Google已经告诉我们这是在所难免的事情。新生的、存在主义的恐慌源于真实与虚拟世界之间模糊的边界（Beane等人）；我们现下所畏惧的是无法识别真实世界与数字景象的边界。

在这样的背景之下，人们会认为对于真实世界的本体性进行探索有些冒失，但是这在逻辑上显得可笑的探索却再一次地播下了一种微妙暗示，多即是好——任何的数据片段至少都有一点点描绘现实的能力。这种流行的观点告诉我们所有的数据中都蕴含着重要的故事，于是，分析与描述数据的工作被简单地等同于发现这些故事。也因此人们更加奉行

数据绝对论，这促成了一种倾向，人们在并不存在规律的地方寻找规律——我们在大量静态的、所谓的大数据的基础上识别和构建出了错误的模式。

在城市的情境下，上述内容有何意义？它告诉我们，在本质上城市数据中蕴含着强大的力量，但它并非总是有效；这些数据对人的行为在无数的时间和空间维度上进行了压缩，人们实际的人际关系与他们所表现出来行为之间存在巨大的差异。我们正在试图从这一团混乱中清理出一种合理并且可验证的观念，例如，重新定义我们的建模、模拟和解读数据的过程，使得我们提取数据价值的过程是可被检查验证的。下面这个案例将会阐释我们是如何取得这种微妙的平衡的。

利雅得和**UTS**项目远景

沙特阿拉伯全境快速的经济发展和人口变迁为这个帝国带来了新的挑战和机会。首都利雅得因为爆炸式的增长而引来了特别的关注，那里的发展速度已经急速地超出了交通设施的运载能力。在1987年至1995年之间，开车出行的比例以平均每年9%的速度在增长；利雅得的道路承载能力本就有限，交通车辆的增长令其变得更加紧张。

UTS项目的目标是建设一个开创性的、高度可变的城市交通系统来应对利雅得特有的交通挑战。为了这个目的，项目的基础是要利用人们日常生活中留下的数字信息来建立模型，对人们的交通行为进行分析、调节和规划，为决策者、规划者、专业的建设者以及利雅得的居民建立一个与一般的智能交通系统不同的交通系统。

从概念上来看，这个项目包含短期与长期两个组成部分。短期内，它的目标是分析人们在现有的交通设施下的交通行为；找到其中所包含的效率低下的部分和城市交通系统中潜在的可改进点。长期愿景则是配合未来的城市发展，预测服务分配和发展的情况，为满足该地区对于交通的高度需求提供助力。

这一项目最高目标的实现有赖于一个移动数据读取设备的构建；一

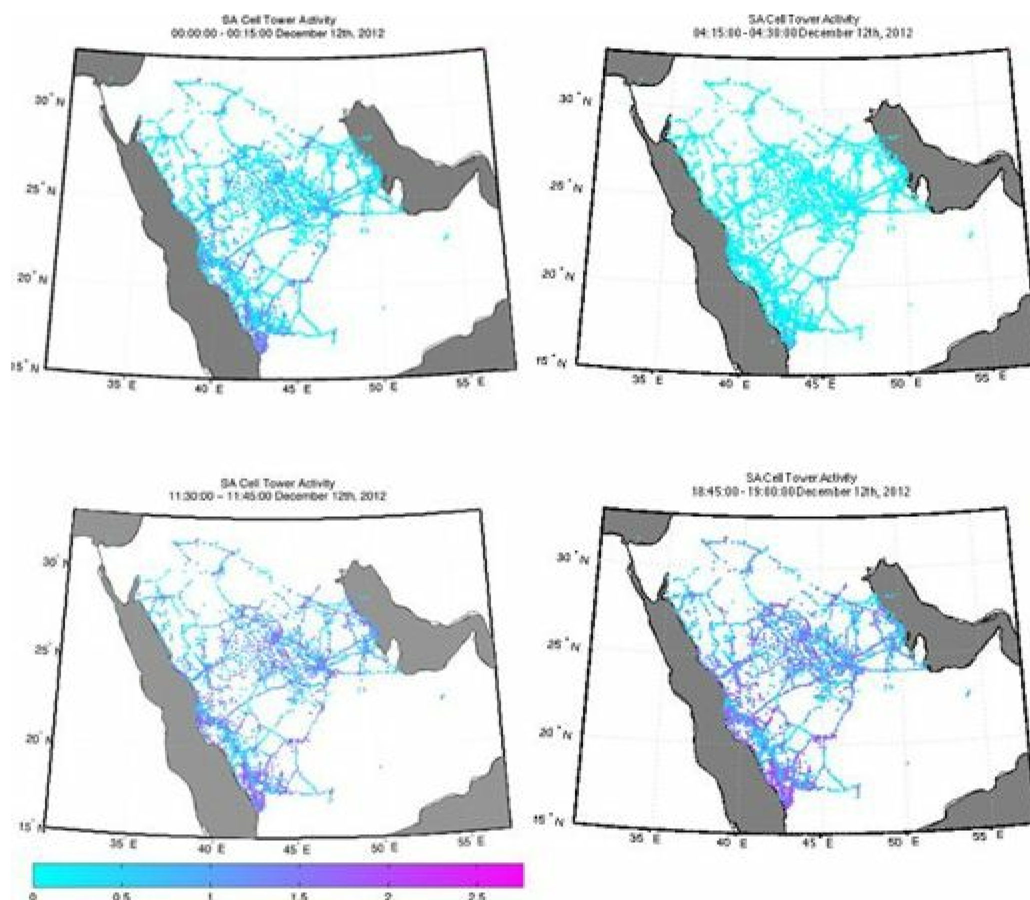
个能够在各种时间维度上便捷地获取、分析、建模和解读多种城市数据集的系统。除了特别关注移动数据之外，工具本身没有其他的通用性的要求，当然，我们希望系统能够有助于对各种通常被视为是异构数据的数据集进行创造性的整合，以发现此前无法感知或者无法明确的关联。

在深入到各种分析的方法论和具体的策略当中之前，我们有必要先明确这个项目的数据基础。我们与当地的电信公司合作，收集到了大约一个月的完整的国内通话行为。对我们来说，移动电话是当前可得的最为有效的实时感知工具之一；数字设备的普及使得我们可以在非常广泛的维度上取得非常明细的人们的行为数据。沙特阿拉伯的移动电话渗透率超过198％——这个令人惊讶的数据表明该国的许多人拥有一部以上的移动设备。我们整合了超过一万个不同信号发射塔的近亿条日常通话。每一条匿名的通话明细记录（Call Detail Record，CDR）都包括该次通讯发生的详细时间和时长、通讯者的位置、通讯类型（电话、短信、网络查询等）以及用户的服务类型（订阅、预付费等）。

分析与可视化：数据中的利雅得



【图1】基于利雅得地形的通话行为的平均数据。图中的高度和颜色代表15分钟内整合的行为。



【图2】利雅得和沙特阿拉伯通话行为快照

我们从将数据空间化和在一系列不同的物理和时间维度上生成行为档案开始。上图展示了一天之内移动电话的应用模式，分别是（1）利雅得和（2）沙特阿拉伯全境。每一个电话发射塔的移动通话行为（电话数量、文本以及数据查询）被聚合起来，每15分钟绘制一幅彩图（从深至浅按对数标度绘制）。这些静态图像为该地区社交生活的动态提供了鲜明的映像（此外，图中也标出了这一地区的通讯主干网），但是基本上它们所提供的只是建议，而不是答案，图像也不能展现人们生活的节奏。城市的影像、基础设施的结构以及时间维度，把更多的隐含信息带到我们面前——我们看到了关于这个城市的生活节奏的更加丰富的表现。

以上的可视化过程在利雅得及其扩展城区的每个像素上，通过颜色、透明度以及高度（仍然是取对数标度）来展示无线通话行为。我们没有将这座城市的无线发射塔当做孤立的点，而是在一个100X100的网格上展示通过插值法计算的网络通信量。在这种方法里，每个网格单元都被分配了一个使用无线电话的活跃度的强度数值，这个数值基于网格与周边无线信号塔的距离，并且经过高斯平滑函数的处理。时间轴上的行为差值也进行了类似的处理，虽然不能通过图像来证实这一点，但是最终的图像呈现出了一个自然得多的景象——对于这个城市的社交特性的令人信服的描绘。以包含卫星图像在内的地图为基础，我们得到了如何通过建筑的形式来表达这座城市的社交节奏的独特影像。作为对于我们数据的一项基础的合理性检查，就像人们所期待地那样，我们发现中心城区在早上交通拥堵爆发之前的时段交通量非常低，在那之后交通拥堵会逐渐向外扩散到整个地区并且在一天中余下的时段里一直持续下去。我们也清晰地看到了数个次要的拥堵中心的出现，它们主要与建筑物的密度有关，而且这些次要的拥堵中心看起来是被公路网络本身所分隔出来的。

这个城市不断变化的交通行为概况也明显地呈现出整个利雅得的交通情况具有非常明显的随时间变动的特性。我们在查看行为的地域分布的变化情况时，发现一个非常独特的特征——我们发现这座城市在正午之前并不活跃，总体来说人们行为的高峰期出现在下午6点15分左右。我们细致地查看，并且综合各个时段的情况发现了一些微妙的地区轮廓线；市中心西南和东北的居民区开始活动的时间比其他地区早很多，并且在一天之内各个时段之间的波动也最为剧烈。最后，时间上的变动在一天之内出现了数个特殊的断点——在这些特殊的时段里，所有的手机通话几乎突然减半。（稍后我们还会提到这个现象。）

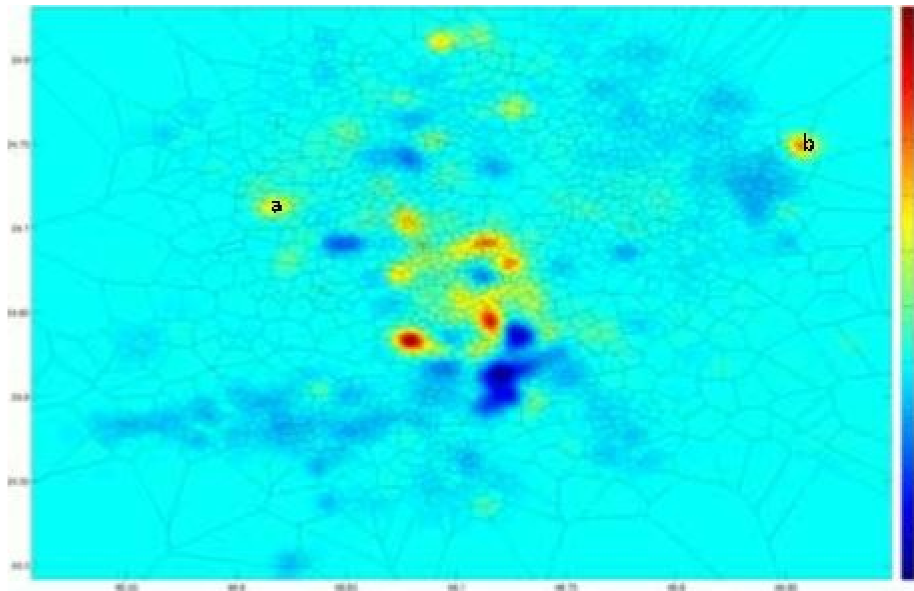
推断人们的居家／工作地点

我们将时间间隔延长来捕捉更为广泛的日夜之间的变化，我们可以

通过这样的区分来识别城市中的居家和工作场所——从方法论上来说，这是推断城市地块用途的前导步骤。我们将那些在工作日晚间被抵达次数最多的手机信号塔周边地区定义为居家场所，而工作日日间被抵达次数最多的信号塔周边地区定义为工作场所。这就要求我们对手机用户进行过滤，找到那些在晚10点至早6点之间拨出的夜间电话大部分（60％）都发生在同一地点、并且在早9点至下午3点之间拨出的大部分日间电话发生在另一个地点的人们。

通过这个过程，我们找到了大约200万个工作日居家－工作场所的配对，于是问题变成了：我们要如何使用这些数据？这些配对信息能够告诉我们一些有关于城市运转结构的新信息吗？当然，每一个居家－工作场所配对在本质上都表明了一段通勤行程的起止点，这使得我们完成的这些工作成为了了解交通需求的基本步骤，而我们是否能够在这些数据中发现一些独特的东西呢？

我们将城市区域内的居家和工作地点整合，并且在地理位置上对数据进行了平滑处理（在以下两张地图中分别展示）。这些图形非常粗线条地展示了在两种不同维度上的土地使用情况。于是我们额外绘制了一张地图来突出显示那些极端的点，图3展示了我们对于工作场所和居家场所的区分。



【图3】工作／居家场所分区地图

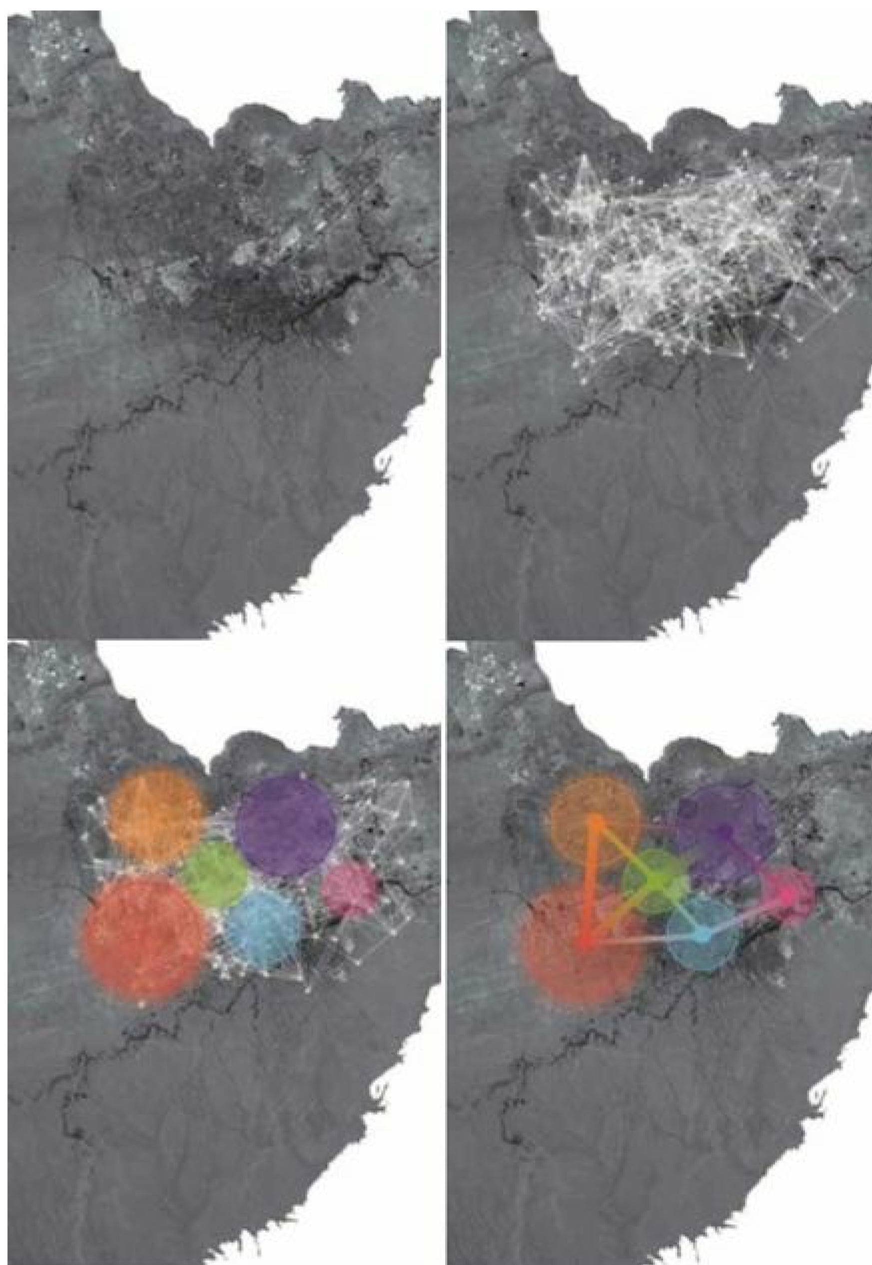
这张地图重点展示了全体居住区与全部工作场所的差异，我们看到一些单一功能的工作热点区域聚集在一起，看来全城范围内都是如此。在城市外围，我们也看到一些大学，它们也呈现强烈的工作场所特征。最后，我们看到城市南侧和东侧人口高度集聚的居民区，期间穿插着小块的空白地区。这与我们查看google地图以及与当地人讨论时所得到的直觉印象再次互相印证。

发现移动的群体

上述的居家／工作场所图像指明了城市组织的逻辑。经验分析证实了我们之前在地图中看到的明显的区域聚集现象吗？如果是的，我们如何形象化地发现其中隐含着的潜在的社群结构？如果我们将所有的居家／工作场所通勤定义为城市内部的移动网络，我们就能使用圈定地区轮廓的方法令人信服地将这个网络分解为子网络。

这个过程起始于全市范围内各移动电话信息塔所在地组成的网络，我们将这个网络中两个节点之间直接相连的权重定义为它们之间累计的通讯流量。这个方法中使用了模块化的优化方案，例如同一个网络中的计算机以一种能够把内部互联中断的总量减到最小的方式组织起来。结

果中的每一个子群落都代表了一个通勤者居住和工作的主要地区。我们在利雅得一共定义了17个不同的群落。



【图4】按照经纬度绘制的拆分结果图

我们将得到的结果与城市的地图叠合在一起，发现了很多有趣的关系。其中最引人注目的是，这些群落都与城市交通主干道有密切的关联。社群的交通行为看起来对道路网络本身有非常严重的依赖，这一点强化了城市对于高速公路基本建设的全面依赖。这个结果也支持了我们

通常所持有的一个观念，那就是交通拥堵的街道在很大程度上制造了隔离和限制，更乐观一点的说法是：好的道路能带来好的邻居。



【图5】可视化图片截屏



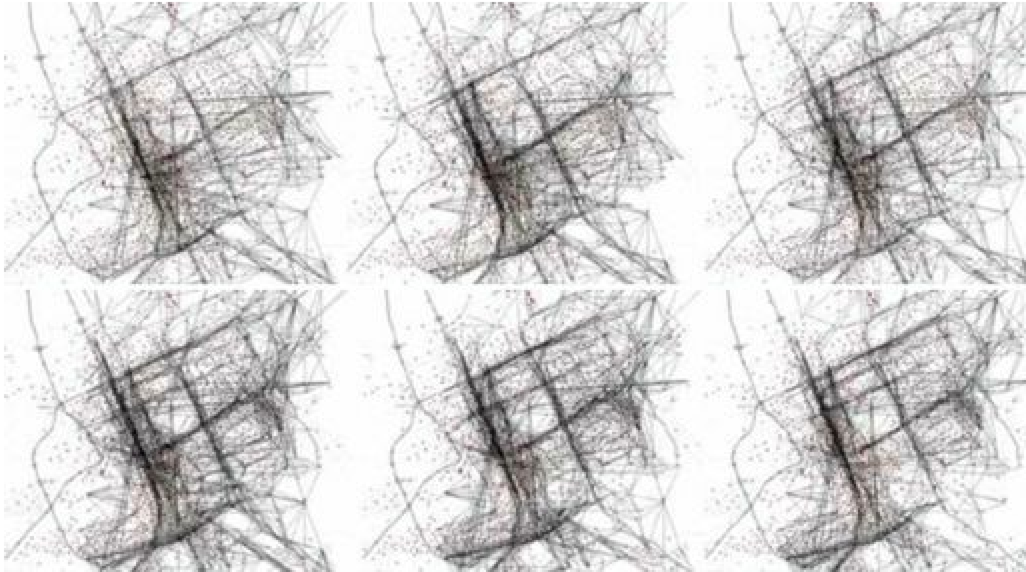
【图6】绘制于道路网络之上的社群详情

从社会节奏到方向性

CDR可以近乎实时地告诉我们城市居民的故事。从规划的角度来看，我们能够收集到的最有意义的故事之一就是个体级别的交通模式，这些信息汇总起来，描绘出了城市分析中最为重要的方面之一：起点／终点矩阵（Origin/Destination matrices，OD矩阵）。构建精确的OD矩阵是交通网络优化中至关重要的一步——这不仅仅是为了评估当下的运力约束，也是为了预测未来的需求。在传统的方式之下，我们通过每5～10年进行一次的、繁冗的人口普查来获得OD信息。这种方法等待时间漫长、耗资巨大，完成之后，只能对于交通需求提供一个基本的快照描述。

有些人提议通过全面安装传感器网络来绕过上述方法的弱点，而我们的方法利用了众多已然存在于我们的生活环境中的传感设备，我们指的是移动电话。我们收集和过滤每位用户使用移动电话的行为，将它们排列成一个手机信号塔所在地点的行为序列，通过这样的办法我们可以评估一个群体中个体出行时包含起止点的交通需求。我们已经证明这种近似的OD流与通过人口普查得到的估计高度相关（Calabrese等人，2006），而且这个办法还有额外的好处，它可以在各种动态的时间切片上捕捉交通需求，从季节变化到小时之间的变化。这样高时间分辨率的解决方案具有重塑我们对于城市交通理解的潜力。

我们首先在小时的维度上构建了OD矩阵，我们希望将其与网络的行为一起呈现出来。在可视化结果的时候，我们将每一段“旅程”表示为一条从起始位置信号塔到终止位置信号塔的弧线。每段弧线都对应着不同数量的行程，为了表示这个特征，我们会根据沿线的行为密度来调整线的粗细（作对数变换）。为了进一步凸显方向性我们使用了配色方案，以蓝色表示起点，以绿色表示终点。绘制这些表示OD的弧线时要求它们在地理上处于同一城市，我们将它们绘制在上述的社会互动网络结构图之上，来试图揭示两个数据集之间那些未被发现的联系。



【图7】利雅得的OD矩阵（早6点至9点，早9点至12点，午间12点至下午3点，下午3点至6点，下午6点至9点，晚9点至凌晨）（译注：原文为12pm，疑有误）

产出的动态地图与当地人对于车流的直觉具有惊人的相似性，而汇总的OD弧线则与下方的街道网络非常相似。非常显著的是，可视化的结果表明沿着城市的主干道有密集的交通行为，如法赫德国王大道以及北侧和东侧的环路。此外，在与社区领袖及规划专业人士进行过一系列研讨会与面谈之后，我们发现这些结果与市民对于通勤模式的主观印象非常相似。为了进一步验证我们的发现，我们将估计出来的OD流量与能够取得的最好的道路统计设备的数据相对比：交通量观测的历史数据。这些交通量数据是通过数个在全城移动的小型气压管传感设备收集到的，一个路口接一个路口，以48小时为周期。

这些计数被以可视化的方式在每个路口的位置以半球标示。每个球体的形状和颜色根据监测到的交通量以每小时一次的频率变换。在城市主干道周边的数据再次与OD数据完美契合，然而我们通过OD数据看到城市中心的东南方向有些非常有趣的行为，遗憾的是我们没有在车辆数量数据中找到可以对照的数据。这是一个未来可以继续探索的问题。

这一系列分析的最后一步，是将这些中等规模的通勤流数据转化为道路网络本身的行为。通过概率变换，我们将收集到的OD交通数据转

化为具体的地理信息系统（GIS）中的路段数据库中的信息，我们有可能对各种交通出行人群对于整体运输网络的影响进行量化——这表明建设一个细致全面的实时通勤展示系统是可能的。这项技术有能力对超载的主干道、启动整个城市日常拥堵的特定地点进行量化。类似的研究已经发现，只是来自少数地区的少数司机造成了主要道路的拥堵。对于王等人根据人们对于本已脆弱的城市道路的依赖，为波士顿市的交通问题识别出了最重要15个普查区（共计750个）。这一点是否在利雅得也是适用的，尚待确认。

思考的前沿：从利雅得看数据

前述研究利用了本质上具有“社交性”的数据来发现在城市中大规模移动的人口的行为特征，结果得到了能够精确地反映利雅得当前的带有方向性的人流变化和循环的结果。在这个探索性的分析之后，我们可以将研究的焦点转移到导言部分中所提到的另一个问题——通过对城市内部人际网络的互动进行过滤发现隐藏的社会趋势。有众多的情境因素使得利雅得，以及整个的沙特阿拉伯王国，成为地理空间文化分析的理想对象。这需要回答许多其他理论问题：城市的结构会激发哪些我们在数据中所捕捉到的社会文化因素？回到约翰·斯诺（John Snow）的霍乱地图，我们是否可以通过将我们的数据与城市结构结合在一起，来发现社会生活中潜藏的方面？以及最终它是否可以更进一步地告诉我们沙特阿拉伯城市结构的特性和构成？下列所述只是进一步研究的思路。它的首要目标是强调那些可以在社会和自然地理模糊的交界处找到答案的问题。这些简单的探索性的思考无论如何也算不上详尽，我们的陈述只是希望能够为未来的研究大致描画出一个范围。

一旦开始广泛接触沙特阿拉伯的环境，我们不可能不注意到这里存在非常明显的性别隔离现象，这是开始进行任何比较社会研究之前所必须了解的。这个国家以其严格的性格规范而著称。阿拉伯女性面临着从强制佩戴尼卡布（面纱）到上街必须有男性陪护等各种措施——这些措

施即使不是对她们无条件地限制，也是过度保护。这种城市生活的背景信息，会对于定义日常生活起到什么作用吗？我们将匿名的CDR与人口统计信息结合在一起，这使得我们可以通过经验来发现这种隔离制度是如何对这个国家产生影响的。特别的是，沙特阿拉伯有一条令人感到疑惑的规则，禁止女性驾驶汽车，这在全球来说都绝无仅有。由于可供选择的公交手段有限，女性在出行时不得不依赖她们的男性亲属或者雇佣司机。我们可以在空间上对这条规定进行量化——我们可以重现和表达这种空间上的限制吗？这结果看起来将会是怎样的呢？哪些区域是女性最容易和最不可能达到的？她们在众多的城市中是如何分布的？此外，我们能否把这种由文化所施加的限制与因城市设计而产生的限制区分开来？这种分隔会对人们的互动网络产生什么影响——空间约束是否在这里仍然起作用？最后，这会为政策制定者、规划者和管理者提供什么启示？

阿拉伯世界另一个独特的文化现象是每日的祈祷。就像之前已经说到的那样，我们在移动电话通话行为中发现了一个有趣的模式，它与我们之前分析过的任何国家和城市都不同。人们会多次中止日常活动，在30到40分钟之后才恢复如常。这些不活跃的“低谷时期”就是每日祈祷的时间。一日五次，全国成千上万的穆斯林放下他们的电话，转向麦加圣城的方向祈祷。店铺和商家理论上来说要停业大约20～30分钟，这期间宗教警察（the Mutaween）就在街上巡查，他们会将所有闲逛的人送到最近的清真寺里去。让我们惊讶的是，我们的行为分布与这一行为非常契合。召唤祈祷的具体时间与太阳在空中的位置有关，因此通过区分CDR在西部、中部和东部地区的分布情况，我们就可以看到如上面图示所展示的全国祈祷时间的变化。这引发了其他一系列有趣的问题。我们已经在上述的“社交脉搏”图示中看到，我们可以从地理图像上识别出这种无线电话通信行为的突发式暂停，但是我们能够对这种变化的强度进行量化并且将它绘制出来吗？我们能够找到祈祷召唤对于哪些区域影响

最大吗？这种暂停与城市环境的空间组织形式之间有什么关系——我们对于强度的衡量与土地利用模式有关吗？它与清真寺的密集程度有关吗？沿着这个思路，我们能够发现并且展示出人们的移动会因祈祷时间导致的通讯暂停而有何变化吗？在临近祈祷时间窗口的时候，出行的平均长度会变短吗？最后，中止强度可以用来表征一个地区的虔诚程度吗？我们能够在虔诚度与人际网络互动之间发现任何关联吗？

我们会继续对这些思考中的问题进行研究，城市的文化脉络将会在寻找答案的过程中成为为我们指明关注方向的核心要素。在这个细致往复的过程中，我们在空间的框架之下和之外都进行了研究，我们希望能够建立一系列模型，来找到研究城市与生活在城市中的各种社会群体的新方法。

参考文献

1．Beane, Silas R., Zohreh Davoudi, and Martin J. Savage. 2012. "Constraints on the Universe as a Numerical Simulation." arXiv Preprint arXiv:1210.1847. http://arxiv.org/abs/1210.1847 Letzter Zugriff 4. Mai 2013.

2．Calabrese, Francesco, Giusy Di Lorenzo, Liang Liu, and Carlo Ratti. 2011. "Estimating Origin-Destination Flows Using Mobile Phone Location Data." IEEE Pervasive Computing 10 (4), pp.36-44.

3．Corner, James. 1999. The Agency of Mapping: Speculation, Critique and Invention." Mappings. 1st ed. London: Reaktion. pp.213-52.

4．Perry, Tony. 2007. "Navy to Mask Barracks Shaped Like a Swastika." Los Angeles Times, September 26. http://articles.latimes.com/2007/sep/26/local/me-swastika26. (Letzter Zugriff 30. April 2013)

5．Tufte, Edward. 2001. The Visual Display of Quantitative Information. 2nd ed. Cheshire: Graphics Press. 24.

6．Wang, Pu, Timothy Hunter, Alexandre M. Bayen, Katja Schechtner,

and Marta C. Gonzalez. 2012. "Understanding Road Usage Patterns in Urban Areas." Scientific Reports 2.

注释

[1] 例如在公交系统中产生的电子车票的数据，其主要目的是为了保证乘客可以利用有效的车票，来登入交通系统。

第三章　空间——设计的意义
建筑环境网络

作者　安德烈斯·塞文随克（Andres Sevtsuk）

译者　姚琪

“城市正因为其组织的复杂性而会出现问题，”对于1960年代，曼哈顿的哈德逊街上的日常生活中的“芭蕾舞曲”，简·雅各布斯（Jane Jacobs）是这样评论的，“变化繁多，但并不是杂乱无章；它们之间相互关联，自成一个有机整体。”（Jacobs, 1969: 433）。那时雅各布斯正在研究了当时流行的城市设计理论，该理论认为通过正规干预可以使复杂的社会生活变得有序。雅各布斯认为，问题在于，对发生在密集的城市环境中社会和经济的相互作用，城市建筑师和规划师缺乏理解，因此造成他们所实行的干预与一个地方真正的需求相脱节。进一步说，完全不清楚环境形态在他们干预成功中所起到的作用，因为众多的文化、历史和地理因素塑造了哈德逊街的活动方式，远不止于形式因素。

“某种程度上说，一个公园的使用程度取决于公园自身的设计，”雅各布斯说，“但即使公园的设计会部分影响公园的使用，反过来，取决于谁在什么时候使用公园，而这反过来又取决于公园周边的城市使用。进一步来说，这些使用因素对公园的影响，其中一部分来自每种因素独立于其他使用因素对公园的影响；另一部分来自与其他因素组合一起对公园作用的共同影响，对于某些组合来说，会刺激彼此互相影响的程度……无论你想尝试对其做些什么，城市的一个公园表现出来的是一个组织复杂性的问题，这就是问题所在。”（Jacobs，1961:433）

50年过去了，描述和分析建筑环境中的复杂空间相互作用仍然是城市设计所面临的主要挑战之一（Batty，2007）。如果以为城市设计师缺乏兴趣或意愿去深入研究社会组织和塑造当代城市的无形力量，会显得有些天真。可恰恰相反，有足够的证据表明，详细地调查多样和复合的

城市环境的工作原理很受设计师欢迎（例如，Belanger等人，2001；Rienets、Siegler等人，2009；Busquets 2006; Sorkin 2009）。

一些评论家认为，城市设计师在社会科学领域内受到的有限教育阻碍了对交互的社会化进程与城市形态进行进一步理解。很多城市社会学家已经提醒城市设计师，要警惕韦伯（Webber）所说的：“一些根深蒂固的教义只是在一些简单的映射模式下寻找秩序，可事实上它潜藏在极其复杂的社会组织中”（Webber 1963）。事实上，虽然大多数设计师没有在社会分析的定性和定量的方法上受过训练但在城市设计领域内外，仍然有着丰富的文献，提供了优秀而严谨的社会分析案例，这些案例涵盖了空间的物理配置和其占据模式之间的复杂交互作用（Gehl，2010；Whyte，1980；Peatti，1968；Gans，1962）。研究表明，对于外行来说可能很复杂的东西，通常隐藏着有待发现的秩序，或如雅各布斯说描述的，有组织的复杂性。

在本章，我们认为，管理复杂城市环境的空间和社会分析，挑战不仅仅来自于用来描述和研究空间形式和功能之间的交互作用的研究方法和分析技巧，还来自于描述研究的问题所用空间表示的规范。我们认为，描述建筑环境最普遍的手段——平面图——带有一定的局限性，使得其难以被用于研究某个街道不同用户之间复杂的空间交互作用。每种建筑环境包含了一种空间秩序，决定了不同建筑、公共空间和连接它们的路线之间的近邻和链接关系。这些关系影响了不同行走路线的使用，公共空间的明显程度或连接程度，或建筑彼此之间交通便捷程度。反过来，这些空间模式决定了对于特定的土地用途来说，选址结果的好坏；不同楼内的住户经常会在哪些公共空间相遇；以及一个空间的活动会对其他空间造成什么样的影响。我们认为，建筑环境的网络表示，为采集和利用这些城市形态的关系提供了一种行之有效的框架。

平面图——是一种二维的描述形式，有时候描述的是建筑环境的某些功能——在城市设计师和学者中仍然是最著名、最有用的空间表现媒

介（Conzen，1960；Moudon，1986；Anderson，1993）。平面图是一种传达空间信息的强有力工具，所采用的表达方式使跨学科的专业人士更容易理解。然而，平面图可能会被误解，而且其所包含的种类丰富的内容和意义很容易被忽略（Mandelbaum，1990；Hoch，2002；Ryan，2011）。也许，研究复杂城市环境最重要的问题是，平面图中所存储的关于建筑环境的大量信息，但是丢弃了在各种元素之间近邻、链接以及互相连通所有这类相互关系，这些关系都需要靠读者的眼睛去测量和解读。平面图并不体现图中元素（如街道、建筑、机构等等）之间明确的关联信息；这些连接需要通过可视化的方法去估计，观察谁与谁进行了互联、如何连接以及为什么连接。另外，平面图拥有丰富的建筑环境元素，但是缺乏传达元素间相互关系的描述；这些分析结果的质量很大程度取决于分析师的素质。

从平面图中读取空间关系是可能的，但劳心劳力，而且远非易事。而且，从平面图中通常容易读取一对一的关系——读取从地铁站到某个特定建筑的路线相对简单——但以相同方式读取一对多的关系就变得极其复杂。计算距离地铁站步行10分钟路程内，所有可能到达建筑以及可用的步行路线并非易事，需要花不少时间。然而，这可能是决定新车站位置的一个重要指标。增加十字路口的约束（例如交通信号灯，地下通道），只关注在某种特定使用类别的建筑上（例如居民住宅）和不同规模的建筑（例如，每栋建筑中居住单元的数量），很快我们就找到了一个复杂到难以消化的问题。企业主根据他们的客户或供应商来选取位置，居民则根据附近的便利设施做出选择，而城市基础设施投资更可能在使用更多的场所得到批准。如果规划的环境是为了吸引所期望的用户，规划的公共空间是为了所期待的活动模式而准备，那么了解这些关系就非常重要。从平面图中衡量建筑环境对这些决策的影响程度很困难，而且需要同时读取空间关系的多种关系集。而更具挑战性的是做这些事能足够的快，跟上某个城市设计师思维过程的步伐。人类的大脑倾

向于串行方式运作，而不擅长同时处理多个并行计算（Minsky，1988）。例如，读者可能会尝试同时去记忆两个或三个有限的数字集合。在实际的城市环境中，空间关系的分析可能需要并行处理成百上千条这样的关系。

为了表示和分析这些复杂空间关系，城市设计师和规划师已经开始使用基于网络模型的建筑环境。与传统的平面图不同，基于网络表示的城市空间对网络元素之间的显性关系进行了编码，例如，记录了一条街道与另一个街道是如何连接的；不同区域、建筑、或房间之间行走需要花多长时间；或者有多少人在这些地方出入。这样的链接信息通常用两种方法中的一种存储。第一种方法，信息可以存储在一个完整的起讫点（O-D）矩阵，其中每列数据表示了平面图中的所有元素（例如，区域、街道、建筑、公司等）附近的全部目的地；而单独的一列用来指示每个这样的连接所需的链接信息。这种链接信息列包含了任意类型的链接信息，如出行时间，出发地和目的地之间上下班工人的数量，在两地之间经济投入或产出的流通总量，等等。这种方法采用了数据库查询，分析相对容易，可以检索出一系列出发地和目的地之间所需的空间关系。但这种便利是以信息存储为代价——用来表示每个位置对之间关系需要一张表格中单独的一行，而所有的位置对就需要很多的表格，以观察地点数目的平方关系增加。对于只有100个地点的情况来说，连接的数量是10,000。如果所有的关系都是对称的，也就是说，从A到B的连接特点与从B到A的相同，那么表格可以减小一半。但是对于成千上万的地点，表格仍然太大以致很难分析。

第二种方法更为经济，即用邻接矩阵表示所有的空间关系。邻接矩阵并不是汇总了环境中每个相关位置对之间完整的路线信息，而是只存储了每个位置直接的邻接关系。如果将环境建模成一个邻近网络，那么邻接矩阵就会采集每个相邻位置之间直接的邻接关系。如果将环境建模成一个建筑和街道的网络，那么邻接矩阵就会捕获沿街道网络上每个邻

近建筑之间直接的邻接关系。实用的网络分析算法可以查询这些信息，并且可以从这张更短的表中推测出网络中所有元素之间完整的空间关系。比起完整的O-D表，查询邻接矩阵需要更加高级的算法，但所需存储空间会大大缩小。如今处理这种表格的算法允许大量的空间交互分析，几秒钟就能得到结果（Vanegas等人，2009）。

用于表示网络和表格中的这些信息有很多不同的方法。然而，重要的不是网络表示所用形式的精确程度——而是围绕在土地使用或城市形态上（Bhat等人，2000），使用实际网络线路或连接的直线距离（Anselin，1988），原始或双重网络表示法（Hillier，1996；Porta等人，2005），双元素或三要素网络（Sevtsuk，2010）——但事实上，在特定环境中的空间关系是用数值描述的，所有这些位置之间的所需的连接都在一个关系表中进行了明确的编码。这些空间关系可以描述交通、材料、信息或者金融交易方面的连接。这与传统的平面图有很大不同，而在过去10年中，这种方式已经悄无声被大多数城市设计者和规划师所采用。平面图的读者并不需要用可视化和直观化的方式去推断嵌入在环境中的复杂空间关系，因为基于网络的表示明确地对这些信息进行了编码，并且允许用户在运行中访问空间连接的大量组合概要。网络模型会自动分析城市空间中大量的并行关系，并允许分析师即时地将这些信息应用在城市设计的决策中。这深刻地改变了我们描述和分析复杂城市环境的方式，为实际规划问题制定更明智的决策铺平了道路。

接下来，我们将描述这些模型中的其中一种——在城市形态实验室开发的城市网络分析工具箱——（Sevtsuk和Meknoonnen，2012）。用于描述建筑环境的有很多其他基于网络的方法；我们使用了一种自己开发的方法来阐述城市空间网络中更普遍的功能。（Levin，1964；Casalaina和Rittel，1967；Rittel，1970；Tabor，1970；March和Steadman，1971；Hillier，1996；Porta等人，2005；Xie和Levinson，2007；Okabe和Sugihara，2012；Miller和Wu，2000；Jiang和

Claramunt，2002；Peponis和Bafna，2008；Vanegas等，2012）。

城市网络分析工具箱——是地理信息系统软件（ArcGIS）中一个开源免费的插件——用三种基本元素对建筑环境进行建模：边缘，表示旅客可以行走的路线；节点，表示两个或以上的边缘相交的十字路口；以及建筑，表示人流在街道与室内环境之间进出的地方。建筑可以由网络中的其他任意位置点替代：公共空间、公交站点、公用设施等。因此，我们要分析的单元变成了某个建筑（或网络中的其他位置标识符），使得可以分别计算每个建筑的相互关系。

每一建筑、街道和十字路口都带有一种额外的属性集，用于描述其现实生活中的属性。这些属性存储在另一个表中，可描述这些元素任意可测量的属性：对于建筑来说，它们的尺寸、高度、构成的混杂度、居住的人口等等；对于街道来说，它们的走向、交通容量、人行道特征等。互联元素的权重表示为研究城市街道网络中建筑之间不同类型的空间关系，提供了一系列的可能性。这种网络表示框架如图1所示。左图是位于马萨诸塞州，剑桥的哈佛广场上一部分区域。在平面图中，用颜色表示土地用途。右图同样是平面图，展示了一个网络。网络中的每个建筑在一个离散的位置——在该案例中是建筑的入口大门——与最近的行走路线相连接。但要注意，一个建筑可以有多个门，与网络有多种连接。



ORIGIN	DESTINATION	DISTANCE	TRANSACTIONS
1	2	242	3
1	3	11	0
1	4	95	0
1	5	141	1
2	1	56	2
2	6	136	8
2	7	25	0
3	9	12	0
3	11	85	0

【图1】左图：哈佛广场的平面图表示。右图：哈佛广场的网络表现，下面附带一个邻接矩阵

某个区域形态和功能的网络表示，为复杂空间分析提供了基础。城市形态的三种元素描述了城市基础设施的物理模式——建筑形态和其行走路线的二维和三维几何表现形式，公共空间的形状，以及连接它们的路径。使用这些类别的属性允许我们进一步区分这些元素的参数——建筑的体积，有关路线中心附近建筑的间距或位置，路线的容量或走向等等。表格属性也允许我们描述每个元素的各种功能——有哪些活动在什么地方发生，有多少人参与了这些活动，以及活动如何彼此连接的。通常可以将活动按照松散型的分组归类，例如生活、工作、或游乐场所，但活动也可以从一种活动转变到另一种，或在每天的某段时间或每周的某些天增强了使用程度。综上所述，这些指标聚合成了某个地方的复杂描述，其中的所有东西可与它周围一切相关联（Tobler，1970）。这些关系不是杂乱无章的，由分析师组织对其进行了明确地编码，并存入邻接矩阵和属性列表中。现在来看看新加坡的白沙浮地区，将这种类型的表示应用在一个真实的复杂城市环境中。

白沙浮位于新加坡市区，约有一平方公里。它是一个旧城区，曾经作为莱佛士计划的一部分开发起来的，该地区坐落了从19世纪到20世纪最初几十年的传统商铺。自1960年代以来，该地区已经逐渐用多层大进深平面的商业结构重建，以适应大规模的、各种类型混合的活动。

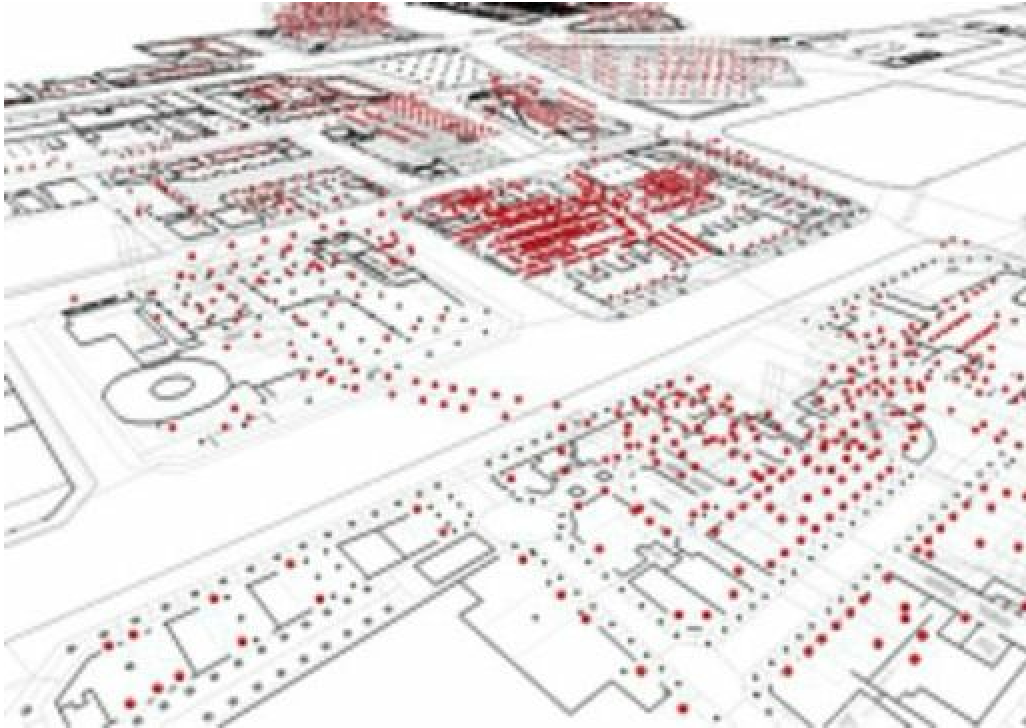
图2表示了白沙浮街的内景图，该区域中心有着一个多层的集市，里面有着数以百计的小型零售和食品商店。总计有超过4,000个个体工商户，包括1,769家零售商，559个服务提供商，519家餐饮场所，130间办公室，38家酒店，24个教育机构，以及19个游乐设施，这些都坐落在一个约0.8平方公里的地方，附近是白沙浮大众捷运（MRT）站。白沙浮是新加坡最忙碌的地方之一，确实有着新加坡最复杂的城市环境。



【图2】新加坡白沙浮街的小巷实景

2012年秋，城市形态实验室的研究人员调查了该区域，详细记录了每扇门、每个建筑和每种商业，以及他们的规模、用途类别和一些其他经济特征。调查涵盖了该区域中所有可以公共出入的楼层，大约有一半的商业区在底楼，另一半在楼上或地下。研究人员还记录了该区域的整个人行通道网络，包括了室内和室外，并且区分了地上和地下——在不到1平方公里的地方，观察了超过32公里的线性步行路线：35％的路线位于室外，26％位于室外但有覆盖物（例如拱廊），以及37％位于不同楼层的室内。

图3表示了网络中这种信息的编码。红点表示个体工商户，灰色线条表示人行通道，黑色线条表示底楼建筑的墙结构。



【图3】新加坡白沙浮地区的三维网络示意图

我们展示的该地区的网络分析使用了两种类型的空间连接指数：中间状态（Betweenness）和可及范围（Reach）（Sevtsuk和Mekonnen，2012）。第一种类型涉及该区域不同位置的行人交通预测；第二种则是对餐饮机构的便利性进行了建模。

为了估计在该地区的不同部分行人在哪里行走以及如何行走，我们观察了整个地区的捷运站到零售目的地的行走路线。根据现场采访，白沙浮的很大一部分人群搭乘捷运来这里购物的。我们的分析基于这样一个假设：从白沙浮捷运站到不同区域的1769个零售商店中的其中一个商店，走的是最短路线。我们使用了UNA工具箱中的中间状态分析，对这些路线进行了建模，并且记录了哪些网段在行走过程中最拥挤。因此，中间状态指标采集了在每个网段大致行人数量，这些人都是沿着最短路线从捷运站到达某个零售目的地。

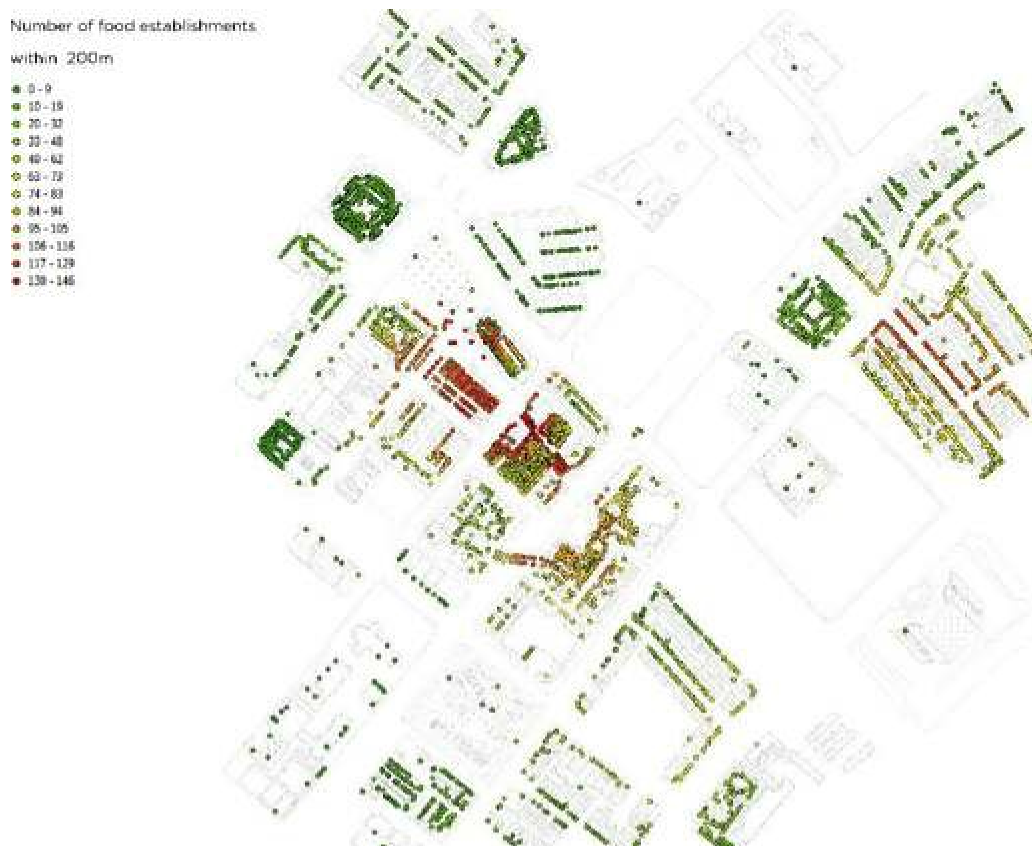
图4显示了分析结果，用颜色对客流进行了编码，从绿色到红色依次变化来表征流量的逐渐增加。我们发现流量最高的步行街活动是在阿尔

伯特广场、白沙浮街和白沙浮枢纽站——这些是该地区中所有主要的购物目的地。在阿拉伯街和哈吉巷也存在一个流量高峰，都是有历史的街道，街道两侧几乎都是老店面。这些区域确实特别拥挤（图2）。或许更重要的是，网络分析不仅仅让我们能够预测一般区域的活动中心，甚至还可以预测出一些特定的路段或贯穿了整个区域的室内走廊有可能吸引更多客流。不同地段的客流移动变化对于解释商业混合和使用模式都是一个重要的因素。例如，餐饮场所常常倾向坐落在其他出发地和目的地之间有较多行人的路线上。不全部依赖行人，还依赖从其他商户流出的有价值客户的商户，可能占据着第二优越的位置，仍然靠近客流，但坐落在街的两边，租金更低的地方（Sevtsuk，2010）。



【图4】中间状态分析，表示了在该地区中从白沙浮捷运站到所有单个零售目的地预期的客流。

当分析该区域中哪些地方到餐饮机构有最佳行走路线时，我们看到，餐馆和酒馆会坐落在与零售目的地非常近的地方，而且步行的路线可以直接导向它们。图5表明，一个基于网络的可及范围指标指定了200米步行半径之内的餐饮机构。可及范围指数估计了有多少特定类型的目的地——在案例中指的是餐饮机构——在出发地开始到给定的步行半径之内（例如，200米）。在给定范围内餐饮目的地越多，指数越高。

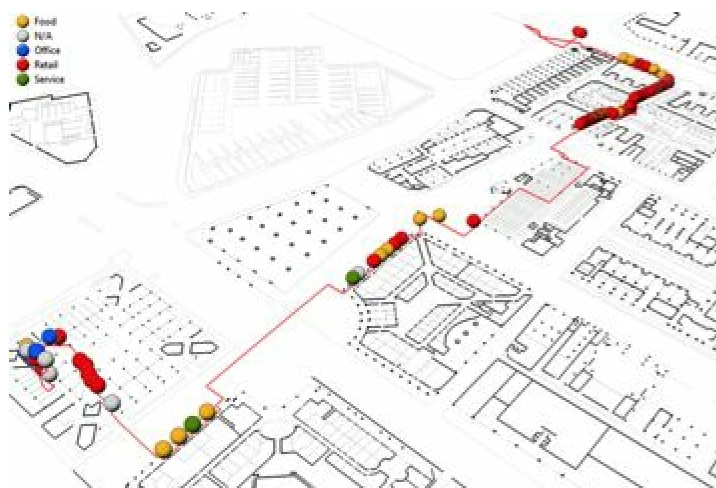


【图5】可及范围分析，表示从每个门出发，在200米步行半径内，可以到达有多少餐饮店。

图5显示的结果表明，餐馆、摊档和酒吧通常聚集在零售商店和直接导向这些商店的步行线路的附近[1]。密集程度最高的是在艾伯特广场和白沙浮街，那里的艾伯特市场和食品中心聚集着众多的食品摊位。在白沙浮街、白沙浮广场，以及地图右侧的岔路阿拉伯街和马士吉街，都有一个餐饮机构聚集区。雅柏中心前面，步行3分钟就能到达146家不同的餐饮店。这4个区域重心都是在食物上面。总的来说，在该地区有3倍的零售食品店，但即使是在地图上淡绿色的位置，200米步行半径范围内也有30到50家餐饮场所，这意味着该地区不缺乏就餐的选择。

如果我们将白沙浮捷运站与某一零售商店——例如森林商业中心电子商城4楼的一家计算机硬件商店——之间的行走路线放大那么在分析中，我们可以进一步定性分析某一特定路线的特点。这是一名普通顾客在该地区逛街经常走的路线，期间会路过86家商店，其中58家是零售商

店，20家是餐饮店，8家办公室和3家服务机构。对比这些路线会导向不同的社会经济目的，比较结果对于很多实际场合很有价值——例如，解释，不同路线对于行人的吸引程度，不同人群对于该城市的体验差异，或者用来研究在机构间的微观经济聚类。



【图6】从森林商业中心电子商城4楼某家商店到白沙浮捷运站，一条典型的步行路线。

建筑环境的网络表示，为描述和分析复杂城市环境提供了一种强有力的框架。该方法已经使用在大量的数字化城市模型中，而且此方法的应用在未来的几年中会迅速增长。不像传统的平面图，城市空间的网络模型明确地对它们表示的不同角色和位置之间的链接信息进行了编码，使得不同环境元素间的复杂空间分析结果可能经过几秒钟的计算机运行，就能得到结果。这些模型的使用，克服了在传统平面图中读取空间关系时缓慢且具有挑战性的处理过程。但因此，空间网络分析会依赖编码进列表中的各种关系。这些关系的记录是采用这些方法重要的第一步。

然而，城市环境的网络表示并不能取代传统的平面图，而是成为它的一种补充。我们知道，作为城市设计师，通过视觉读取平面图会无与伦比的细致、敏感而且强大。平面图读取对于城市规划专家来仍将是至关重要的。嵌入在这些平面图底层的基本网络连接可以简单地使用强有力的空间互联来丰富静态表示，否则这种空间互联很难被度量。它们有

助于实现费时费力的计算和测量任务的自动化，这样，平面图的读者也许就不需要费劲脑汁，而且还能同时利用这些信息来研究或处理这张平面图。

网络模型的图形规划接口也将使分析师能够克服一个过度参数化的模型所具有的缺点。当然，网络模型体现出来的形态和功能的相互关系并不很全面，而且会缺失某个地方很多重要的维度——该地方的历史或更广泛的社会、文化或环境背景。但是，网络模型的图形接口能够用传统的平面图一样的方法，去解释这些内容，以作补充。城市空间分析的一种整体方法仍然需要“一种包含了背景的务实看法，并且能够在不同观点中寻求连续性”（Hoch，2000: 54）。平面图元素之间的数字互联改善了而非妨碍了整体思考。

最后，我们应该记住，某个地方的一种新型表示未必能够帮助进一步地理解其底层的复杂性。但通过提供一种清晰的框架，描述嵌入在架构中多种同时存在的空间关系，建筑环境的网络模型能够消除可视化读取那些关系的主要负担，并且允许设计者专注于分析本身，而不是问题的描述。

参考文献

1．Anderson, Stanford. 1993. "Savannah and the Issue of Precedent: City Plan as Resource," in Settlements in the Americas: Cross-Cultural Perspectives, edited by Ralph Bennett. Newark, DE: University of Delaware Press, 110-144.

2．Anselin, Luc. 1988. Spatial Econometrics: Methods and Models. Dordrecht: Kluwer Academic Publishers.

3．Batty, Michael. 2005. Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals. Cambridge, MA.: The MIT Press.

4．Bélanger, Pierre, Chuihua Judy Chung, Joshua Comaroff, Michael

Cosmas, Sonal Gandhi, A. David Hamilton, Lan-Ying Ip, Jeannie Kim, Gullivar Shepard, Reshma Singh, Nathaniel Slayton, James Stone, and Sameh Wahba. 2001. "Lagos: Harvard Project on the City," in Mutations, edited by Rem Koolhaas et al. Barcelona: ACTAR.

5. Bhat, Chandra, Susan Handy, Kara Kockelman, Hani Mahmassani, Anand Gopal, Issam Srour, and Lisa Weston. 2000. "Development of an Urban Accessibility Index: Literature Review." Center for Transportation Research, The University of Texas at Austin. Report completed for the Texas Department of Transportation.

6. Busquets, Joan. ed. 2006. Aleppo. Rehabilitation of the Old City. Cambridge, MA: Harvard University Graduate School of Design.

7. Casalaina, Vince, and Horst W. J. Rittel. 1967. "Generating Floor Plans from Adjacency Matrices."

8. Conzen, M. R. G. 1960. Alnwick, Northumberland: A Study in Town Plan Analysis, 2nd rev. ed. (1969). London: Institute of British Geographers.

9. Gans, Herbert J. 1962. The Urban Villagers: Group and Class in the Life of Italian-Americans. New York: Free Press of Glencoe.

10. Gehl, Jan. 2010. Cities for People. Washington, DC: Island Press.

11. Hillier, Bill. 1996. Space Is the Machine: A Configurational Theory of Architecture. Cambridge: Cambridge University Press.

12. Hoch, Charles J. 2002. "Evaluating Plans Pragmatically," Planning Theory 1, no. 1 (March): 53-75.

13. Jacobs, Jane. 1961. The Death and Life of Great American Cities. New York: Random House.

14. Jiang, Bin, Christophe Claramunt, and Michael Batty. 1999. "Geometric Accessibility and Geographic Information: Extending Desktop GIS to Space Syntax," Computers, Environment and Urban Systems 23, no.

2: 127-146.

15．Levin, Peter Hirsch. 1964. "The Use of Graphs to Decide the Optimum Layout of Buildings," Architects' Journal 140, no. 15: 809-815.

16．Mandelbaum, Seymour. 1990. "Reading Plans," Journal of the American Planning Association 56, no. 3 (Summer): 350-356.

17．March, Lionel, and Philip Steadman. 1971. The Geometry of Environment: An Introduction to Spatial Organization in Design. London: RIBA Publications.

18．Miller, Harvey J., and Yi-hwa Wu. 2000. "GIS Software for Measuring Space-Time Accessibility in Transportation Planning and Analysis," GeoInformatica 4, no. 2: 141-159.

19．Minsky, Marvin. 1988. The Society of Mind. New York: Simon & Schuster.

20．Moudon, Anne Vernez. 1986. Built for Change: Neighborhood Architecture in San Francisco. Cambridge, MA: The MIT Press.

21．Okabe, Atsuyuki, and Kokichi Sugihara. 2012. Spatial Analysis Along Networks: Statistical and Computational Methods. Statistics in Practice. Hoboken, NJ: Wiley.

22．Peattie, Lisa Redfield. 1968. The View from the Barrio. Ann Arbor: University of Michigan Press.

23．Peponis, John, Sonit Bafna, and Zongyu Zhang. 2008. "Connectivity of Streets: Reach and Directional Distance," Environment and Planning B: Planning and Design, 35, no. 5: 881-901.

24．Porta, Sergio, Paolo Crucitti, and Vito Latora. 2005. "The Network Analysis of Urban Streets: A Primal Approach," Environment and Planning B: Planning and Design, 33, no. 5: 705-725.

25．Reilly, William J. 1929. Methods for the Study of Retail

Relationship. Austin: Texas University Bureau of Business Research monograph no. 4.

26. Rieniets, Tim, Jennifer Sigler, and Kees Christiaanse, eds. 2009. Open City: Designing Coexistence. Amsterdam: SUN.

27. Rittel, Horst W. J. 1970. "Theories of Cell Configuration," in Emerging Methods in Environmental Design and Planning, edited by Gary T. Moore. Cambridge, MA: The MIT Press.

28. Ryan, Brent D. 2011. "Reading through a Plan." Journal of the American Planning Association 77, no. 4: 309-327.

29. Sevtsuk, Andres. 2010. "Path and Place: A Study of Urban Geometry and Retail Activity in Cambridge and Somerville, MA," Ph. D. diss., MIT, Cambridge, MA.

30. Sevtsuk, Andres, and Michael Mekonnen. 2012. "Urban Network Analysis: A New Toolbox for ArcGIS," Revue international de géomatique 22, no. 2: 287-305.

31. Sorkin, Michael. 2009. Twenty Minutes in Manhattan. London: Reaktion Books.

32. Tabor, Philip. 1970. Traffic in Buildings 4. Evaluation of Routes. Land Use and Built Form Studies. Working paper 20. Cambridge: University of Cambridge.

33. Tobler, Waldo R. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region," Economic Geography 46, no. 2: 234-240.

34. Vanegas, Carlos A., Daniel G. Aliaga, Bedřich Benes, and Paul Waddell. 2009. "Interactive Design of Urban Spaces Using Geometrical and Behavioral Modeling," ACM Transactions on Graphics 28, no. 5 (December): article 111.

35. Vanegas, Carlos A., Ignacio Garcia-Dorado, Daniel G. Aliaga,

Bedřich Benes, and Paul Waddell. 2012. "Inverse Design of Urban Procedural Models," ACM Transactions on Graphics 31, no. 6 (November): article 168.

36. Webber, M., 1963. "Order in Diversity: Community without Propinquity," in Cities and Space: The Future Use of Urban Land, edited by Lowden Wingo. Baltimore: The Johns Hopkins University Press, 23-54.

37. Whyte, William Hollingsworth. 1980. The Social Life of Small Urban Spaces. Washington, DC: Conservation Foundation.

38. Xie, Feng, and David Levinson. 2007. "Measuring the Structure of Road Networks," Geographical Analysis 39July): 336-356.

我们的城市在多大程度上是多中心的？

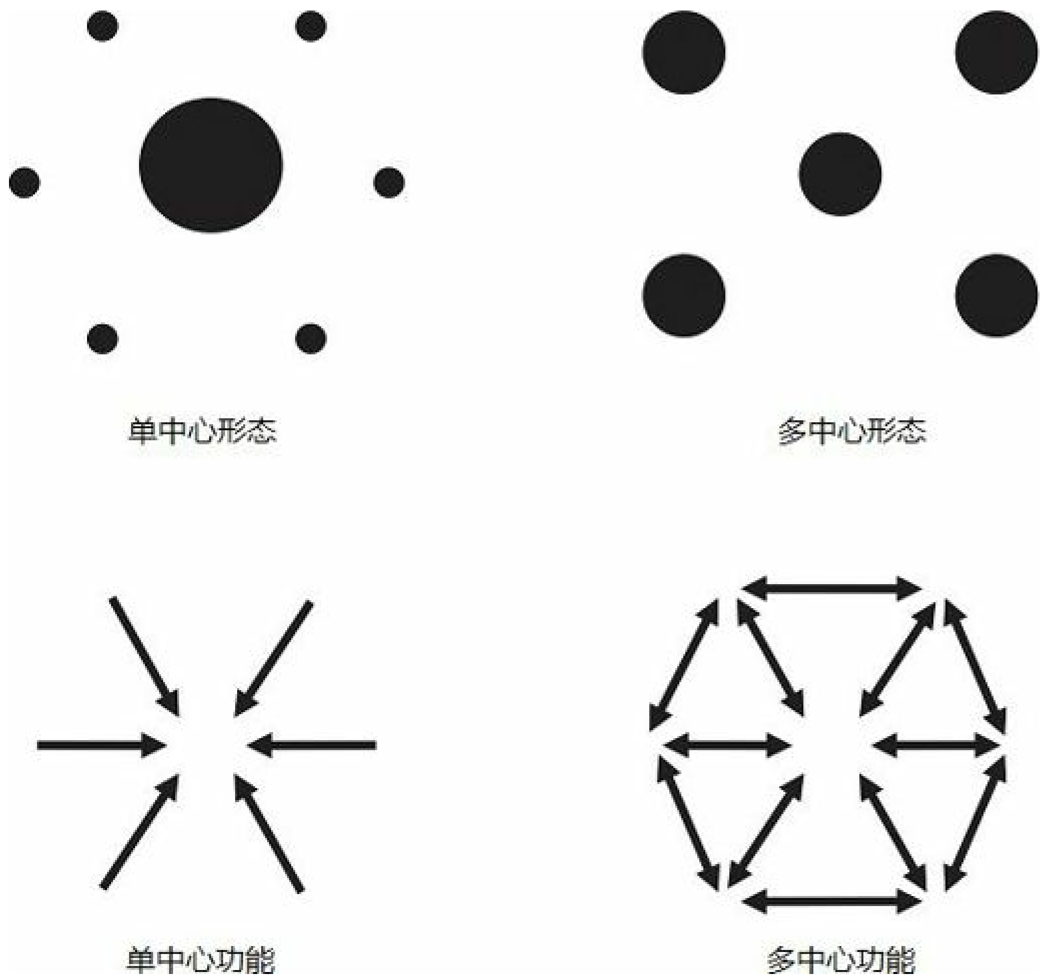
作者　马库斯·施莱费尔（Markus Schläpfer）

译者　党霄羽

如今的大都市，从纽约和伦敦到新加坡和东京，都是由许多相互联系的中心组成的：那些用于工作、娱乐、休闲和交通等日常活动的地区都是高度密集的。对城市空间结构进行深入的、定量的了解对于当代城市来说十分重要，这些地区未必总是那么明显地具有中心区的特征。例如，预测新的商业中心如何和已经存在的中心进行互动，并且最终改变了全市的交通模式，对于预先防范交通拥堵和空气污染是一个先在条件。最近许多新的、大规模的人类活动的数据，例如从移动电话网络中自动搜集的数据，为日益增加的多中心共存带来了前所未有的巨大潜力，同时为城市规划提供了数不胜数的实施方案、基础设施投资，并减少对环境的影响。

和城市中心以及城市的分中心有关的城市景观的空间结构，是随着时间的推移不断发展的。从历史的角度来看，私人汽车拥有量的爆炸式增长，以及公共交通的大量增加，带来了上个世纪几乎所有主要城市的转型：从一个非常简单的单中心结构，转变为了复杂的多中心结构。前者的中心商业区（CBD）与周围的郊区泾渭分明，而后者有多个共存的活跃的中心（Anas等人，1998）。这一趋势在近些年来还将持续，为我们带来提高城市性能的新的挑战。因此，近年来在城市规划与经济社会领域对城市多中心的关注不足为奇。然而，由于其多方面的特点，多中心或者多中心主义仍然是一个相当模糊的概念，缺乏一个普遍为人所接收的定义。一般来说，如图1所示，多中心有如下两个重要特点：（1）形态尺度，即中心的面积与空间分布；（2）运作规模，即另外处理了不同的中心之间的联系，诸如日常通勤的人流量，或者是商业网络与社会网络之间的联系。（Green，2007；Burger和Meijers，2012）。传统

上，采用人口、就业或商业密度的简单阈值这样一种相当直接的方式，评估多中心形态。然而，对功能性多中心进行恰当的量化却是困难的，主要是因为缺乏大规模的、基于个人行为或者是社会互动的数据。因此，多中心已经超出了这些基本的概念，我们对城市多中心与分中心下的人们的时空互动知之甚少。



【图1】多中心的形态和功能规模，来自Burger和Meijers（2012）

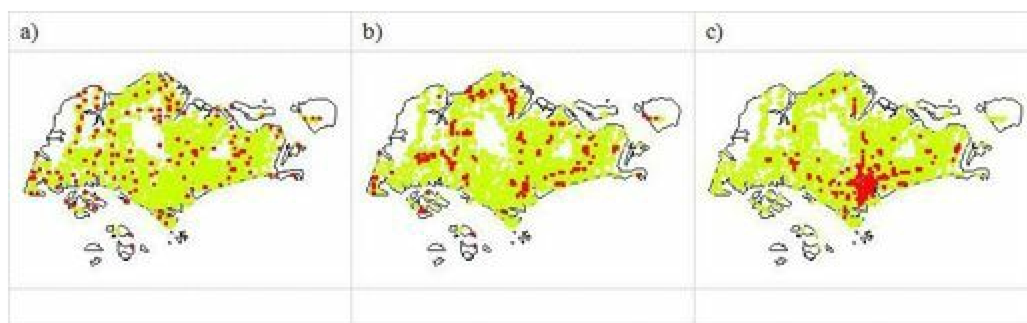
在这些前提下，用户从移动手机、信用卡、在线社交网络与许多其他来源产生的数据集的新浪潮（见第二章），给我们提供了前所未有的时机，让我们可以仔细观察并最终量化这些城市内在的互动。最近的研究已经通过基于Oyster卡的电子票务系统，分析了伦敦数以百万计的地铁交通人次，展示了在车站与人流方面的不同层次的中心，从而在这个

方向迈出了第一步（Roth等人，2011）。为了更好地利用城市居民的电子足迹，下一节提供了一个应用方面的例子，分析了大众移动电话的呼叫数据记录（CDR），用以对多中心的大城市进行量化。CDR为我们提供了基于单独个体在城市内的移动绘制的精确统计图（Isaacman，2010），由于移动电话普及率的提高，这个数据覆盖了很高比例的人口。显而易见的是，这些实际做法比传统的基于调查问卷的调查或者普查数据要优越，后者采集更加困难，费用更高，只能局限于小部分人口，而且仅提供一种静态的快照而已。

在一项正在进行的、由可感城市实验室发起和领导的项目中，我们正在开发新的算法和数据挖掘技术，让我们能够把CDR转化为人们的时空分布，从而进一步发现复杂的多中心模式（Schlapfer等人，2014）。在SMART中，新加坡这个城市国家的最大的移动电话运营商之一为我们提供了以学术研究为目的的大量的CDR。近期以计费目的收集了大量数据，涵盖了一半以上的人口，连续几个月，涉及每天近1千万人次的通话记录。每条记录对应着一次单独的通话。其中包括（1）两个移动电话使用者之间的匿名号码（例如使用代理）；（2）通话时长；（3）通话起始的时间戳；（4）两基站的呼叫路由线路，并包含它们的地理位置。

这个参照地理位置的电话数据通过识别电话使用者们在城市内日常活动的集中，来识别那些最重要的位置，从而揭示新加坡的多中心的结构。我们分析过程的第一步是通过所谓的起始（O-D）矩阵，这一概念来自于交通工程领域。简而言之，我们的算法通过傍晚或者夜间通过信号塔拨打的电话，为每个移动手机估计了起始位置（或者说来源地）。然后我们把新加坡的整个城区划分为0.5公里×0.5公里的网格（也就是说包括至少1423个基站），并确定每个网格内每个用户电话打往何处（或者称为目的地）。我们分析的第二步，是计算出每个网格内生活在特定距离之内、以特定频率访问特定位置的用户数量。结果得出了一个具有

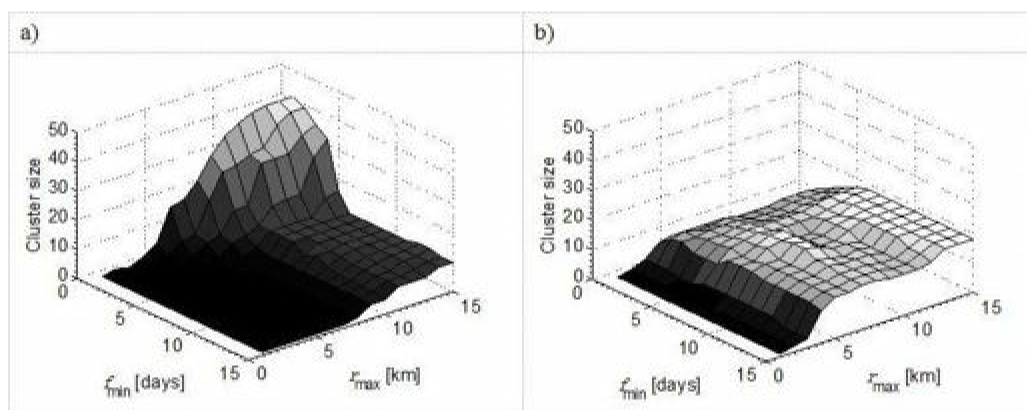
相对重要性，或者说是吸引力的综合图片，网格中的每个像素，告诉我们被吸引的人次、频率以及他们和中心的距离。反过来，这使得我们现在能以全市的视野系统地研究中心的出现。例如，如果我们只考虑小半径范围内的吸引力，把我们的关注点限制在那些只在家附近活动的出行者，我们就会发现这些吸引大部分人的中心散落在整个城市区域内（见图2）。这种模式是直观的，因为它能够突出那些当地的购物设施和其他设施这样具有吸引力的地点。然而，如果我们通过考虑更多的住得更远的出行者，来逐步扩大吸引力的半径，那些能够吸引大量人群的网格在空间上就变得越来越集中，它的确恰好对应了那些我们主观感知上认为是城市中心的地方。举例来说，把每个网格半径扩大为15公里左右，我们可以直观地识别出新加坡的著名的城市中心，诸如城市南部的核心区域、西部的裕廊、以及北部的伍德兰德（Woodlands）（见图2b）。如果我们进一步扩大每个网格的面积，包括了从整个城市来的出行人员，我们就可以进一步观察到包括已经存在的中心区在内的多中心结构，正如图2c所证明的那样。因此，新加坡的城市中心吸引了城市中大多数的居民，而裕廊等其他中心则在更小的区域内吸引了大多数的出行者。换句话说，多中心的程度高度依赖于对空间尺度的估计。



【图2】新加坡的时空多中心展示。标记（为黑色）的像素对应着相对人流量最高的10个地点。A）区域内半径（r＝1km）的低值是高度本地化的，并且可以观察到分散的中心。B）对于中等距离（r＝5km）重要的中心区域开始出现。C）对于大型区域的半径（r＝15km）来说，市中心就变得清晰可见。

我们超越了肉眼观察，在一定人口涌入的情况下，根据基站连接

（具有共同的边界）的数量来衡量每个城市中心的规模，并且通过增加空间尺度来量化每个中心相对重要性的变化。如图3所示，以市中心和裕廊为例，这两个中心的大小都随着区域半径的增加先是增加，然而，当超出了某个特定的半径之后，裕廊的大小就稳定下来了，而市中心的人流量还在增加。随着地理规模的不同，不同城市中心的重要性也产生了变化，正如我们在新加坡的例子中观察到的，这意味着城市的不同中心和分中心有内在层次的组织。



【图3】根据相互连接的基站（集群）的数量来衡量在相对人口流入量方面新加坡排名前10％的城市中心的相对重要性，并且使用了区域半径rmax和最小访问频率fmin构建了一个函数。A）市中心；B）裕廊

除了明确空间尺度在多中心特征方面起到的作用之外，CDR还提供了在时间尺度方面可供观察的丰富资料。为了达到这一目的，我们评估了新加坡的每个网格被访问频率所带来的影响。和距离的增加类似，我们只考虑那些以最小频率访问特定区域的用户。我们基于网格中用户访问网络的最小天数，来界定这些用户。图3展示了两个示例中心的相对重要性，也是用中心的规模来表征，是一个最小访问频率的函数。当把范围限制在仅仅对常客进行分析的时候，中心区的相对重要性就急剧较少了。因此，市中心地区的重要性仅仅是非常客带来的。相反，更像是本地中心的裕廊，其相对重要性随着访问频率的增加甚至稍有提升（图3），因而主要吸引了那些定期游客。这表明，城市中心在时间方面的

吸引力，与在空间方面的吸引力是一样重要的。有趣的是，我们基于城市内部的人流的发现，似乎符合经济地理学的长期原理，解释了拥有特定商品和服务的中心地段能够吸引大量的人群，但时间不固定（Krugman，1996）。一个著名的例子是克里斯塔勒的中心地理论，这一理论解释了城市的空间布局（Christaller，1933）。

新加坡目前的研究主要是作为一个例子，根据对大都市组织复杂结构的深层理解，来说明对此前难以获得的用户数据进行分析有巨大前景。特别是，我们利用手机CDR提供的信息，揭示了城市居民的足迹，进而明确了时间和空间尺度这两个城市多中心的要素。这些基本的见解可能是一个起点，会朝向功能性多中心的一种正式的、数据驱动定义迈进。此外，有意思的是，我们能把在不同中心之间的人流加以量化，以及把这个结果跨越不同的地域、文化和经济来扩展到其他主要城市，并且衡量这些中心的在出行行为或者诸如排放量等影响环境的要素，来比较其表现和效率。因此，有必要在尊重隐私的条件下对必要的数据进行搜集，这需要新颖的数据挖掘、存储、匿名化的处理办法（Trantopoulos等人，2011）。此外，希望这些见解能够帮助城市规划者以及决策制定者来把握可持续的城市发展方向，同时保持生活的高品质。

参考文献

1．Anas, Alex, Richard Arnott, and Kenneth A. Small. 1998. "Urban Spatial Structure," Journal of Economic Literature 36: 1426-1464.

2．Burger, Martijn, and Evert Meijers. 2012. "Form Follows Function? Linking Morphological and Functional Polycentricity," Urban Studies 49: 1127-1149.

3．Christaller, Walter. 1933. Central Places in Southern Germany, translated by C. W. Baskin. London: Prentice-Hall, 1966.

4．Green, Nick. 2007. "Functional Polycentricity: A Formal Definition in

Terms of Social Network Analysis," Urban Studies 44: 2077-2103.

5．Krugman, Paul R. 1996. The Self-Organizing Economy. Cambridge, MA: Blackwell Publishers.

6．Roth, Camille, Soong Moon Kang, Michael Batty, and Marc Barthélemy. 2011. "Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows," PLoS ONE 6, no. 1: e15923.

7．Schläpfer, Markus, Michael Szell, and Carlo Ratti. 2014; in preparation. "Human Spatial Organization and the Emergence of Urban Centers."

8．Trantopoulos, Konstantinos, Markus Schläpfer, and Dirk Helbing. 2011. "Toward Sustainability of Complex Urban Systems through Techno-Social Reality Mining," Environmental Science & Technology 45, no. 15: 6231-6232.

城市问题：用复杂系统理论看城市本质的新视角

作者　易斯·贝当古（Luís M. A. Bettencourt）

译者　陈然

城市不是生物体，它们不过是机器，甚至连机器也不是。城市自身不会增长，不会改变，也不会自我复制或自我修复。城市不是自主实体，不以生命周期运行，也不会被感染。［……］但是，更困难、也更重要的是，要看到这一隐喻之下根本的荒谬性，以及它如何引导我们轻率地做出一系列举动：减少贫民窟，以阻止其“传染式的”的传播；寻找城市的最佳规模；反对持续性的增长；区分用途；努力保持绿地；抑制竞争中心的发展；防止“结构混乱的无计划开发”；诸如此类。凯文·林奇，《城市形态》（1984）。（Kevin A. Lynch, Good City Form）

简·雅各布斯挑战

本文标题取自简·雅各布斯（Jane Jacobs）在她颇具影响力的著作《美国大城市的死与生》（Death and Life of Great American Cities，1961）中提出的挑战。她在书的最后一章中把城市定义为有序复杂问题（organized complexity，），为她就城市形态的讨论提供了概念基础。复杂性科学在当时是全新的观点，雅各布斯偶然中接触到瓦伦·韦弗（Warren Weaver），并很快意识到这个概念对于理解城市的重要意义。同样，韦弗在向世界解释克劳德·艾尔伍德·香农（Claude Elwood Shannon）新提出的信息论时，也颇具独创性（Shannon和Weaver，1949）。当一系列新的整合性视角和概念像野火般在许多学科燎原时，便正是科学史上激动人心的时刻。不过，对当时来说，要做到用更精确的术语定义城市是何种复杂系统为时尚早。形式更清晰的社会网络又过了50年才出现。今天，我们在全球城市观察到了许多新的城市形态和增长模式，形式多样、范围广泛，但这些在当时仍属未知。

在我看来，界定城市问题是个不小的挑战，难度远远超出简·雅各布

斯时代在原则上拒绝城市重建规划的做法。它要求的也不仅是考虑和采纳她对城市空间本质的观察结论。在现在看来，雅各布斯的部分结论可能显得有些狭隘，太局限于她上世纪50年代末期对纽约西村（New York's West Village）的观察。今天的难题在于，把城市看作复杂适应系统，使用更现代化的语言，为城市树立全新的、更合适的概念，并使城市规划在这些原则指导下，取得更具科学实践基础的发展。

我的主要目的是，结合50多年来的研究，以及城市作为复杂系统这一新兴科学的新视角，对城市的本质特性展开讨论，解决这个难题。

我会阐明，把城市作为在时间和空间维度内的综合社会网络的视角，以开放式土地利用和基础设施建设的一般特性为前提，这让我们得以建立关于城市化的一个新的统一模型。这一视角在定量上得到了大量实证研究的支持，这些研究描绘了过去和现在全球数千个不同发展水平的城市的特性（Bettencourt，2013）。在城市设计方面，这一城市概念强调了生成模型（generative model）的重要性，即局部结构仍由具备特定目的和信息的主体发展形成，但同时必须受到城市整体功能的限制，相当于一个开放性的“社会反应堆”。

新视角：作为复杂系统的城市

我们已经没有什么借口去忽视城市的可量化特性。现在，全球各地城市的可知性都比以往任何时候更强，在社会、经济、基础设施、空间等多个方面都是如此。

从一系列地理和复杂系统新研究中（Bettencourt，2013；Koolhaas，1995；Batty，2005；Bettencourt等人，2010；Nordbeck，1971；Bettencourt和West，2010）得到的主要实证结果可以简单地总结为：随着城市发展，它们不仅可以实现一定的空间规模经济，同时还能实现一般社会经济生产力的提高（2008年新城市主义大会；Alexander等人，1977）。例如，当我们比较同一城市系统（通常是国家）内两个人口规模不同的城市时，我们可能会发现，人口规模较大的城市更密集，人均

基础设施存量（道路、管道、电缆）更小。我们还会发现，大城市往往生活更富裕（以人均GDP或人均工资衡量），物价更昂贵，在文化和技术上更有成效，比如在创造性尝试或专利数量方面。因此，这些结果都证实了一种人们普遍熟悉的预期，即大城市不仅更拥挤、更昂贵，而且在文化上更有成效、更令人兴奋。

这些影响十分普遍，在许多城市系统内的城市中都可以观察到，从美国到中国，从巴西到德国，都是如此（Bettencourt，2013）。它们也普遍适用于各个时期，不论是当代城市系统，还是已经过两千多年检验的前西班牙时期墨西哥盆地的居住空间格局，这种格局的形成必定也是与同时代其他格局相互独立（Ortman，2013）。更具体地说，人们发现这些城市特性随着城市规模的变化而逐渐变化，但与具体城市规模无关。平均而言，城市人口每增加一倍，人均基础设施存量就减少10％至20％，财富创造速度、创新速度提高10％至20％，并且，如果任由暴力罪行等人类社会经济活动的不良产物自由发展的话，那么它们发生的可能也会上升10％－20％。尽管城市的定义各不相同，但从整体上看，这些结果都是适用的，不论是将城市看作混合种群还是统一的劳动力市场。这些定义带来了一种视角：把城市功能视为大都市区（Bettencourt，2013），而不是核心功能区，或其他没有任何实际社会经济意义的特定行政区。对小城市来说，这种区分不太重要。

这些发现，特别是它们从发达城市到发展中城市的普遍适用性和重要性，当然是一个起点，但仍没有提出关于城市是什么的新模型。为此，我最近提出了一个新的理论框架（Bettencourt，2013）：我为城市设定了一个简单的概念，把它看作同时处在时间和空间维度内的社会和基础设施网络，以定量的方式推导出这些特性，及一些其他特性。

其中心思想是，城市首先是大型社会网络。从这个意义上说，城市不仅是大量人群的集合，也是社会联系的集聚。空间、时间和基础设施在社会互动的形成和存续过程中都起着基础作用，通过增强联系，使这

些社会互动成为开放性的互动，实现在能源利用和人力方面的可持续性。

从这个角度看，我们可以从以下四个简单假设中得出所观察到的结果，即城市随着人口规模增长而表现出的规模特性：

（1）城市使社会网络互相融合，每一个人事实上都与城市中的任何其他人有着联系，即使他处在最小社会关系网中。这决定了人员、商品和信息的流动成本。当其他条件相同时，一个城市可能因为人口规模更大，也／或可能因为流动成本相对较高而更密集。

（2）基础设施网络是分散的，意味着它们是随着城市的发展而逐步建成，新增基础设施数量与特定城市的空间密度成正比。这就是说，不存在一个所有来往都必经的单一中心（没有核心城区）（Samaniego和Moses，2008），基础设施网络也是随着人口规模增长和科技进步而逐步形成。

（3）个人与社会保持联系的努力与城市规模无关。这是有可能的，因为（1）和（2）表明，随着城市发展，它们在空间上更密集，因此，即使个人移动的距离保持相对不变（恒定作用力），他能够联系到的人和机构却在增加。换句话说，随着城市发展，城市来到了我们的身边。这消除了社会学和社会心理学上（Samaniego和Moses，2008；Simmel，1903；Wirth，1938；Milgram，1970）对大城市的重要反对理由，使城市规模基本不受限制。

（4）城市的社会经济产量，包括经济生产力、创造力，以及犯罪行为，都与单位时间内实现的社会交往数量成正比。平均来看，这个数量由公共网络中的人口密度决定。

擅长数学的读者可以在参考文献（4）（Bettencourt，2013）中读到对这些论述更正式的解释，以及下面将会讨论到的一些其他特性的推导过程。

这个城市概念的一般性意义是什么？首先，把城市看作社会网络的

视角强调了加强个人联系和社会融合，从而令城市实现其全部社会经济潜能的重要作用。事实上，由于多种原因的存在（暴力、隔离、缺乏充足的交通工具），一些城市仍只是维持着初步联系，这些城市的经济表现往往不及融合程度更高的城市。有人认为，增强城市内的社会联系的确是促进城市发展和提高公民参与的关键（Lynch，1960），如通过劳动者分工和合作机会的增加，以及更复杂的社会和经济组织的设立（Holston，2008）。还有人指出，以增进联系为目的的城市规划（如通过协调交通和住房政策，）通常比试图单独解决这些问题的规划更成功。当然，就小规模城市来说，这是智能增长模型（models for smart growth，2008年新城市主义大会，）的核心原则，但这些结论强调的是，在规模巨大的大都市区中社会融合的必要性，而不仅是如社区这样的地方层面上。

这个概念在城市土地总体利用方面有几个重要影响。首先，土地价格随人口增长而上涨的速度高于平均收入增速。这是人均密集度和人均经济生产率同时增长的后果，因此每单位面积和单位时间内的支出增加，例如，人口规模每增加一倍，土地租金平均增长50％！正是土地价格的上升间接地产生了一些自发的解决方案，降低了大城市人均能源使用量和碳排放量。汽车停车费用上升，为了使楼面价与收入保持同步，更高的大楼成为必须，因此表面积与体积的比率减小。这反过来又提供了降低人均制热和制冷成本的机会。这些影响还创造了用公共交通代替私家车的可行条件，即使时间成本很高，即使点对点快速交通的需求正在不断增加。因此，矛盾的是，大城市更密集的土地利用和更高的经济生产力反而产生了一种预料外的结果，即大城市可能更加“绿色”。增加人均土地供应和降低交通成本的政策（如城市重建），尽管解决了其他问题，却会降低城市密集度，并且，在建筑和运输中为了实现相同的社会功能会消耗更多能源。

这个概念还提出了人口密度的问题。当我们对同时期内不同规模的

城市加以比较时，我们可能会发现，在同一个国家内，规模较大城市更密集。然而，在美国，以及在越来越多的其他国家城市，我们同样发现了许多现代化的城市形态，尤其是许多低密度的大城市，如亚特兰大或达拉斯。难道这些城市比简·雅各布斯所了解的纽约西村，或智能增长模型规划者主张的可步行城镇更小吗？把城市当作互动网络的视角告诉我们，当交通技术和经济发展水平发生变化时，这些不同的城市形态如何能够共存。城市的总体空间范围由互动性与移动的相对成本之间的互相作用决定。当然，个人对某种社区关系的偏好也起一定作用。一般而言，当有可能在空间内快速移动时，城市往往变得更透明，更可能在保持社会联系的同时扩展空间。事实上，城市快速交通技术（现代城市轨道交通，特别是汽车和道路基础设施）在发展中国家的普及，正在推进它们用地面积的快速增长，增速有时超过人口增长速度（Angel等人，2011）。当然，这可能会造成问题：例如，如果交通成本与收入水平相比突然上升（如因为与石油价格的关系），那么城市可能就无法保持社会联系，导致预期社会经济产出的下降。反之，对于人口规模、经济活动活跃程度等出现下降的“萎缩”城市，如底特律等前工业城市，这个概念也同样适用：通过缩小城市的空间规模，可以以更小的成本实现遍及更广的社会联系，并期望这些地方能增加联系、重现活力，从而“修复”它们的社会结构。

这个理论另一个有趣的影响是，随着城市规模变大，基础设施总量预期增长速度比土地面积更快。这是从基础设施网络是分散的且逐步形成这一设定中得出的巧妙的数学预测（见注释）。在实践中，这意味着大城市基础设施网络总量在其城市空间中所占比例越来越大。这种影响在大型和／或发达城市非常直观清晰，电缆、道路、管道等都成为城市景观的一种常见特点。但最终它们都因缓解拥堵而非其他基础性用途而向第三维延伸，无论是地下基础设施网络，还是地下的基础设施网络。因此，为了促进发展，尤其是大型密集城市的发展，对与城市规模不相

称的土地利用和基础设施覆盖区域的变动做好规划是必要的。

问题的另一个方面是，我们能够预测在以维持城市联系为目的的运输过程中的能源消耗量（不仅包括人的移动，还有物品、信息、服务等的流动）。运输过程中的能源消耗速度高于人口增速，考虑到空间和实体规模经济，这一结果令人诧异，事实上，它表现出的特性就和收入或创新方式一样。这也是经营城市基础设施的预期成本（但不是一般的建筑）：研究人员已经对该成本进行过计算，如德国城市的电网电阻损耗（Bettencourt，2013）。它也可能以大城市交通更拥堵的形式出现，主干道（高速公路）上的交通情况往往比小型道路上更糟（Downs，1962）。这为城市基础设施成本管理问题，以及促成城市可持续性带来了新的解决方案。我们应把投资看作城市规模的一个非线性函数，策略性地做出提高效率的投资，减少运输过程中，尤其是更宽阔区域的城市基础设施中的能源消耗。

此外，该理论还指出了如何使运输过程中的能源消耗最小化的具体做法，以及其中的一些限制。特别是，它告诉我们社会产出与运输中能源损耗的比率，该比率可以衡量城市效率，与城市规模无关。这可能是城市规模千差万别的重要原因之一。大城市的社会经济集聚优势与它们的耗散成本成正比。拥堵成本的上升是社会经济集聚优势的写照。运转良好的城市能最大限度地扩大社会净收益与基础设施成本之间的差距。这些衡量方式为我们提供了一整套指标，指引我们做出合适的城市规划和政策（Bettencourt，2013）。

到这里为止，我已经证明了城市形态和功能如何受到一系列互相联系、从本质上捕捉城市问题特性的原则的决定，至少从总体上看如此。这些理念仍需更多的实证检验和进一步发展，但最近从移动通话技术中得到的（Schläpfer等人，2013），对于人类社会互动直接细节的测试，以及从前西班牙时期墨西哥城市系统中观察到的，与现代城市类似的规模空间特性（Ortman，2013），都进一步证明了这些原则的普遍性。

在我们探索城市对人类生活及国家发展的重要性的过程中，还存在两个更深层的问题。首先是要更直接地了解，人类互动结构如何产生对于研究如城市里的犯罪行为或创新活动的经济学家和社会学家来说已经十分熟悉的影响。近期，通过对从移动通信中得到的个人联系数据的直接测试（Schläpfer等人，2013），研究取得了一些进展，但还远远不够。更深入地了解这个问题也有助于更成功的公共空间规划（Whyte，1980）。第二个问题是城市系统内城市之间的互相作用。对于不同时期的，包括人口、经济等方面在内的城市发展的测试结果，是处于同一个城市系统中的城市共有的特性。但关于它们是如何建立及由什么因素决定的问题，可能是经济学和城市学中最重要的尚待解决的问题之一。但现在，随着搜集不同规模、不同国家、不同时期的数据，我们将对此问题建立起更清晰的视角，新的理论成果可能将会出现。

规划未来

全球都在以惊人的速度经历着城市化，正如印度、非洲和拉丁美洲的城市发展都是自发进行一样，城市化也没有遵循总体建筑规划。今天全球可能有近10亿人住在贫民窟（联合国人居署2003年报告，<United Nations Human Settlements Programme，2003>）还有更多的人在自己建房，自己组成社区，既不借助建筑师和城市规划师的服务，也不具备任何城市科学知识。城市大转型一直都在进行，导致未来几十年内建成的城市基础设施将超过整个人类历史，而且这种城市转变在很大程度上都是无规划的、自发的过程。在这种混乱而重要的城市背景下，城市化和城市科学的作用是什么？

对于那些参与城市规划实践的人来说，城市空间的复杂系统视角可能还有太多未明确的地方。尤其是，它对规划中一些基本的选择只字不提，比如街道、社区、房屋和建筑的形状、空间的具体用途、分区等。

一种更激进的说法是，城市规划应该使这些选择留有不确定性，由当地个人、机构和社区自行发展。然而，城市历史和如何建立与发展复

杂系统的基本科学概念恰恰都表明了这种激进的说法是对的（Hayek，1945；Anderson，1972）。有时这被称为“规划者的问题”。问题取决于大量异质主体组成的复杂系统中信息的基本特性：实践中，规划者无法得知在城市历史进程中人们可能发展城市空间的所有方式。更了解具体信息的主体通常会做出更好的、合乎目标和期望的选择，只要这些选择不限制其他人的类似选择，也不抑制城市的融合。因此，确保一般约束（比如前一节中提到的），加上一些地方级的基本准则，如受多个世纪以来的乡土建筑、或许多新都市主义或衍生设计启发的准则，也许能为城市规划提供一个可行的模式，尤其对于那些很大程度上非正式建立起的城市。把城市看作复杂系统进行规划，从整体而不是局部视角出发进行设计，帮助城市从一般化到独特化进行打造，这一过程最终可能会实现帕特里克·格迪斯（Patrick Geddes）的心愿——“尽可能地减少改变，提高各层次人民的幸福感，不论是社会最底层的人还是金字塔尖的人。”

我希望我已经阐明，在一般情形下，城市是人类社会中自发形成的自然系统，只要它对不同规模的人类社会都存在优势。从这种意义上看，城市和蜂箱或珊瑚礁一样，都是自然存在，不应被当作随意的人工产品，按照人类意愿重新设计。同时，城市这种复杂信息系统与其他受到更多限制的社会组织形式有本质区别，事实上，在一些能够体现和生产的信息形式上，城市这种系统远比后者更复杂。城市同时展现了人性中最好和最差的一面，包括人类的创造力和想象力，也包括暴力或歧视倾向。因为城市对人类发展的这种巨大潜能，它不应被视作应受控制或抑制的系统，而应被鼓励自发地、开放性地发展，展现对人类共同本质最佳的表现形式。

这也是我们面临的挑战。我们正在经历城市大转型的最后几十年，正如大多数物种群体一样最终会实现全部的潜能，我们终将合力创造地球历史上全新的东西。在不远的将来，我们将能实现古老的人类愿望，

如消除极端贫穷，结束严重不平等，人人拥有健康身体，并且，我们将以可持续的方式做到这一切，与地球生态圈保持平衡。所有这些都会在城市里发生，并将会很快成为现实。大数据和更科学的城市研究方法肯定有助于此。但对于我们所有对城市规划产生影响并参与实践的人来说，最终的挑战是转化、应用及进一步发展这些新理念，促进形成以可持续性和开放性的人类发展为目标，能够鼓励并帮助实现全部社会创造力的城市环境形态。

致谢

我要特别感谢Nicholas de Monchaux，我们就城市规划与城市化的历史，及它们与复杂系统概念的关系展开过多次讨论，这有助于我提出在本文中阐述的想法。还要感谢Michael Mehaffy和迪特马尔·奥夫胡贝尔，他们对本文的较早版本给出了很多评论和建议。这项研究部分得到了洛克菲勒基金会、詹姆斯·麦克唐纳基金会（批准号：220020195）、美国国家科学基金会（批准号：103522）、约翰·邓普顿基金会（批准号：15705）的资助，并接受了Bryan J. and June B. Zwan基金会的赠予。

参考文献

1．Alexander, Christopher, Sara Ishikawa, and Murray Silverstein. 1977. A Pattern Language: Towns, Buildings, Construction. New York: Oxford University Press. Anderson, Philip W. 1972. "More Is Different," Science 177, no. 4047: 393-396.

2．Angel, Shlomo, Jason Parent, Daniel L. Civco, Alexander Blei, and David Potere. 2011. "The Dimensions of Global Urban Expansion: Estimates and Projections for All Countries, 2000-2050." Progress in Planning 75, no. 2: 53-107.

3．Batty, Michael. 2005. Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals. Cambridge, MA:

The MIT Press.

4．Bettencourt, Luís M. A. 2013. "The Origins of Scaling in Cities," Science 340, no. 6139: 1438-1441.

5．Bettencourt, Luís M. A., and Geoffrey B. West. 2010. "A Unified Theory of Urban Living," Nature 467, no. 7318: 912-913.

6．Bettencourt Luís M. A, Geoffrey B. West, José Lobo, and Deborah Strumsky. 2010. "Urban Scaling and Its Deviations: Revealing the Structure of Wealth, Innovation and Crime across Cities," PLoS ONE 5, no. 11): e13541.

7．Bettencourt, Luís M. A., Horacio Samaniego, and HyeJin Youn. 2012. "Professional Diversity and the Productivity of Cities," http://arxiv.org/abs/1210.7335 (accessed April 21, 2014).

8．Bettencourt, Luís M. A., José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B. West. 2007. "Growth, Innovation, Scaling, and the Pace of Life in Cities," Proceedings of the National Academy of Sciences 104, no. 17: 7301-7306.

9．Congress of New Urbanism. 2008."Canons of Sustainable Architecture and Urbanism: A Companion to the Charter of New Urbanism,"http://www.cnu.org/canons (accessed April 21, 2014).

10．Downs, Anthony. 1962. "The Law of Peak-hour Expressway Congestion," Traffic Quarterly 16, no. 3: 393-409.

11．Duranton, Gilles, and Matthew A. Turner. 2011. "The Fundamental Law of Road Congestion: Evidence from US Cities," American Economic Review 101, no. 6: 2616-2652.

12．Hayek, Friedrich A. 1945. "The Use of Knowledge in Society," American Economic Review 35, no. 4: 519-530.

13．Holston, James. 2008. Insurgent Citizenship: Disjunctions of

Democracy and Modernity in Brazil. Princeton, NJ: Princeton University Press.

14. Jacobs, Jane. 1961. The Death and Life of Great American Cities. New York: Random House.

15. Koolhaas, Rem. 1995. "What Ever Happened to Urbanism?" in S, M, L, XL, edited by Rem Koolhaas and Bruce Mau. New York: The Monicelli Press, 959-971.

16. Lynch, Kevin. 1960. The Image of the City. Cambridge, MA: The MIT Press.

17. Milgram, Stanley. 1970. "The Experience of Living in Cities," Science 13, no. 3924: 1461-1468.

18. Nordbeck, Stig 1971. "Urban Allometric Growth," Geografiska Annaler. Series B, Human Geography 53, no. 1: 54-67.

19. Ortman, Scott G., Andrew Cabaniss, and Luís M. A. Bettencourt. 2013. "Urban Scaling in Prehispanic Central Mexico." Santa Fe Institute Working Paper 13-01-001, http://www.santafe.edu/media/workingpapers/13-01-001.pdf (accessed April 21, 2014).

20. Samaniego, Horacio, and Melanie E. Moses. 2008. "Cities as Organisms: Allometric Scaling of Urban Road Networks," Journal of Transport and Land Use 1, no. 1: 21-39.

21. Schläpfer, Markus, Luís M. A. Bettencourt, Sébastian Grauwin, Mathias Raschke, Rob Claxton, Zbigniew Smoreda, Geoffrey B. West, and Carlo Ratti. 2012. "The Scaling of Human Interactions with City Size," http://arxiv.org/abs/1210.5215 (accessed April 21, 2014).

22. Shannon, Claude Elwood, and Warren Weaver. 1949. The Mathematical Theory of Communication. Urbana: University of Illinois Press.

23. Simmel, Georg. 1903. "Die Grosstädte und das Geistesleben." Published in English as "The Metropolis and Mental Life," in The Sociology of Georg Simmel, translated and edited by Kurt H. Wolff. New York: The Free Press of Glencoe, 1976, 409-424.

24. United Nations Human Settlements Programme. 2003. The Challenge of Slums: Global Report on Human Settlements. London: Earthscan Publications.

25. Weaver, Warren. 1958. "A Quarter Century in the Natural Sciences," The Rockefeller Foundation Annual Report 1958. New York: The Rockefeller Foundation, 1-122.

26. Whyte, William Hollingsworth. 1980. The Social Life of Small Urban Spaces. Washington, DC: Conservation Foundation.

27. Wirth, Louis. 1938. "Urbanism as a Way of Life," The American Journal of Sociology 44, no. 1: 1-24

28. World Bank. 2013. Planning, Connecting, and Financing Cities - Now: Priorities for City Leaders. Washington, DC: World Bank.

区域分隔的数字化方法

作者　斯坦尼斯拉夫·索博勒夫斯基（Stanislav Sobolevsky）

译者　黄艺

全球化和信息科技已经加快了社会和经济进程，越来越快地改造着城市、地区甚至整个国家。城市化仍然在扩大范围：在19世纪初叶，全球人口只有3％住在城市。在20世纪到来之前，这个比例已经增长到14％。2008年世界城市人口终于突破50％大关[2]，在现今高度发达的国家甚至已经达到60－80％。

农村和城郊人口不仅在城市化的动态发展中起重要作用，同时也受到城市化的影响。城市的通勤者们原本每天在城市和乡村／城郊之间流动，他们的通勤活动使得城市延伸到其行政边界以外，构建了所谓的功能化城市，覆盖了更广更大的都市范围和城市群。

在现代城市、地区和国家中，越来越多的空间和信息演变给城市和区域规划师们制造了新的难题。对地区的定义在几年之内就过时了。然而我们仍应该对城市里实际的空间互动与界限保持一定的理解，这有助于政府设立合适的地区性社会和经济政策，以及作一些基础建设的决定，比如在规划交通系统的时候尽量按照使用者的需求作出最优的规划。发现地区结构的变化能够为社区和社会的变迁提供重要的线索。

在过去，类似的决策只能依靠调查问卷对于人口流动性的估计。调查问卷是人工的，昂贵而且执行缓慢，一般隔几年才进行一次。现在的文化要求我们迅速应对不仅快节奏，而且庞大的城市中各种各样的社会和经济挑战。这种调查问卷的方式已经越来越不切实际了。

然而新挑战也意味着新的解决方法。不同层面的人际互动中更普遍地涉及到电子和数字信息，也意味着越来越多人类活动留下了数字化的印迹。电话网络运营商会储存通话的时间、时长以及呼叫者和接收者的大致位置。用信用卡买咖啡或者食品杂货时也会留下类似的印迹，被信

用卡服务商和银行记录。甚至出租车旅程也会因为车载GPS系统而生成详细的数字化信息。把这些信息集合起来就形成了庞大的人类活动数据集，也就是人们经常讨论的大数据。现在这些信息可以从政府机构和私人公司取得，而且都被匿名化以保护人们的身份信息。在学术研究方面，这些数据为研究人类行为的规律和模式——流动性，时间上的演变，尤其是人与人的互动——开启了前所未有的可能性。我们可以把人与人的互动看作链接不同地区和形成新地区的原动力。

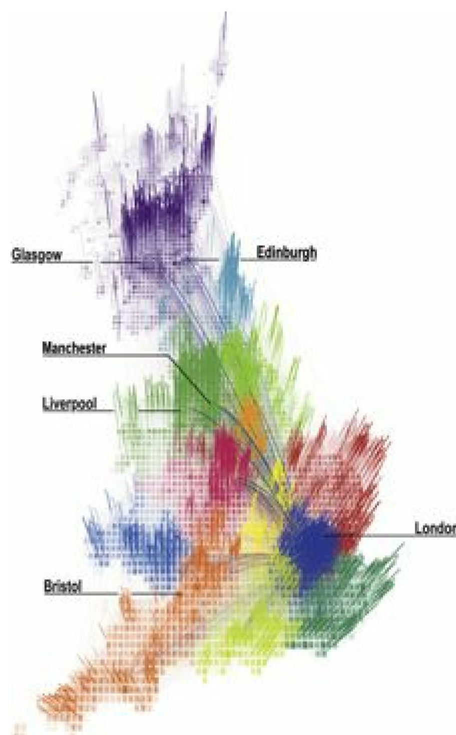
我们利用人类活动的大数据来估量人与人的互动，基于此开发了一个计算方法用以识别地区间的分隔。我们将展示空间上人类的互动如何在社区中体现，这些社区比一个人、群体或社区的社交网络更大，然而仍然是空间上可以描述的社区。“数字化革命”经常会得出这样的结论：数字化的人类互动方式会降低甚至消除物理空间的作用。尼古拉·尼格拉庞特（Nicholas Negroponte）有一个广为人知的预言：“后信息时代将会彻底除去地理上的限制。数字化的生活将会越来越少地依赖于特定的空间和时间”（Negroponte，1995，165）。然而我们将在下文说明空间仍然是定义我们生活和互动的关键因素。

在我们的研究中，我们运用了整个国家范围内人们交流活动的数据记录，也就是电话服务商在一个时间段收集的手机和座机通话。

地理上而言，这些数据包含了多个国家的信息，如美国，英国，法国，葡萄牙，意大利，比利时，象牙海岸和沙特阿拉伯。在最近的一篇文章中（Ratti等，2010）我们关注了英国电话网络的数据。数据覆盖了英国95％以上的住宅和商业座机线路，包括了一个月内的120亿次通话。为保护个人隐私，电话号码在我们收到数据以前就被服务商匿名化。根据地区内电话交换设施的位置，该数据集将每个通话者的地理位置整合在一起，构建了平均面积49平方公里的地区，这个划分粗略得以至于无法获取每个单一用户的地址。

依据这些数据中我们建立了一个大概有2,080万结点，8,580万无方向

链接的网络用以链接有相互关联的用户，即双向打过电话的用户。我们假设以上的网络是在全大不列颠范围内对于个人层面上人际互动的一种估量，并将网络整合为一个有3,042平方像素的网格，每个像素代表9.5公里乘以9.5公里。我们将每个像素当作一个空间上的结点，并且测量该像素与另外每个像素之间连接的强度，就可以导出一个矩阵来描述地理网络中每一对空间结点间的所有双向交流（图1）。这样一来，两个结点间有方向的加权链接就可以被定义为双方互相交流的累计交流量。



【图1】英国的电话互动

到现在为止我们描述的区域划分方法无视了地理上空间的划分；这些方法在最初也并不是用于定义区域的。最初这些方法是基于被称作模块度优化的方法，用于发现网络内潜藏的结构和社区群组（Newman，2003－2006）。模块化这个变量是一个描述潜在社区结构的归一化质量方程，量化了社区内部链接的强度以及社区之间联系的薄弱度。具体来说，模块化测量方法对于社区内每个链接打分，分数是链接实际强度和平均强度的差值，其中平均期望强度基于一个拥有同等总强度，但是链

接强度均匀分布的网络。然后每个社区结构的模块化总分由所有社区内部链接分数的总和算得。

对于电话互动网络而言，这代表我们将两地之间实际的交流量和平均期望值作比较，而平均期望值基于模型内部的假设，也就是假设所有单一链接强度都是全网络的平均值。这样一来，比平均值强的链接就更容易被包含在辨识到的社区内，而薄弱的链接被忽略。

接下来，我们的社区辨识算法迭代地改变社区的初始结构以更好地辨识社区分隔，比如说将两个已有的社区合二为一，拆分已有的社区，或者将一个社区的一部分移到另一个社区，然后再次测验网络模块化总分。上述步骤会一直重复，直到分数再不能提高为止。我们再次重申，这个方法目前为止都没有考虑地理因素，因为它只考虑网络的拓扑结构。最后得到社区的数量也并不是事先决定的，而是由算法在优化过程中选定的。

最后对于英国的划分在图2中可以看到，每个不同区域都有不同的颜色，而11个一级统计用区域命名（Nomenclature of Territorial Units for Statistics）[3]的官方边界由黑色实线标示。



【图2】英国座机电话互动网络的区隔和NUTS2区域的对比

首先令人惊讶的是，由算法推导出的社区在空间上代表了地理连贯

的区域，虽然在计算过程中没有考虑地理因素。另一个令人惊讶的发现是，由模块度优化过程得到的社区数量和官方的区域数量基本一致。

如果比较我们推导出的社区和英国官方行政区隔的具体形状，我们观察到很好的对应性；比如9个英格兰官方区域中的两个——西南英格兰和东英格兰。其它地区比如说苏格兰，看上去和用我们的方法得到的区域一模一样。然而不同处仍然可见。伦敦地区看上去比官方版本更大，然而这并不令人意外，因为主要城市周围的都市地区高速发展，已经超出了已有的界限。另一方面，威尔士分成了北和南威尔士，而中间部分看上去和西米德兰连在一起。以上的结果证实了交通和区域研究领域文献的假设。2001年人口普查的通勤数据也确认了威尔士仍然和它东面的英格兰邻居紧密联系，即使威尔士有其独特的文化和语言传统（Nielsen和Hovgesen，2008）。

在伦敦西侧新区域的发现证实了早先一项关于高新技术活动在“西部萌芽”的研究（Hall等，1987）。伯克郡、白金汉郡和牛津郡总增加值（Gross Value Added）和国家职业证书（National Vocational Qualification）指数显示，这个地理上紧密联系的区域通常在经济活动中表现突出，而且贫困度低[4]。虽然我们仍需要更加详尽严谨的研究来下最后的结论，但是所有以上的观察结果都显示这个模块度优化方法可以基于全国的电话通话，正确地辨识出各个紧密联系的区域。这个归组的方法甚至比官方NUTS的区域界限能更好地表现出一些人与人间的互动。

现在让我们看看上述方法能否适用于英国以外的国家，并且在加入手机通话数据的情况下能否让我们的分析更加准确。

在英吉利海峡对面的法国，我们有幸研究了一个手机网络。数据包含1400万匿名化用户在45天时间段内大概1.2亿次通话，他们的地点也由比手机塔位置来表示，比上文的位置标示更加精细，不仅显示出人的流动性，还显示出人与人的互动。

结果（图3）显示出和官方界限的对应比原先更加清晰，除了一些西南面的局部偏差和其他小的局部偏差。图3以及更进一步的比利时和葡萄牙的研究结果可以在最近发表的文献中找到（Sobolevsky等）。这次跟原来略有不同的分区算法在Blondel等人的文章［7］也有描述。和法国的研究案例不同，比利时手机网络里用户的地点被定义为官方备案的居住地址，而不是通话时的实际地址。即使如此，在由算法选择最优区域数量的情况下，我们看到官方的地区结构和基于电话网络划分的空间分布（图4）有明确的对应性（Blondel等，2010）。如果我们限定最终社区的数量为二（算法可以让我们限定迭代改进的步骤）来计算该国家的最优分割，可以得到一个有意思的结果。在图5中我们清楚地看到比利时分为两个主要部分——法兰德斯和瓦隆，正是沿着荷兰语和法语的分界线划分的。令人惊讶的是，整个网络所有交流的97％都是在这两个区域的内部发生的，而只有3％的交流是在两个区域之间。在一个链接均匀分布的网络中，这个值应该只有50％，这也强调了比利时这两个区域之间有很大的隔阂。另外在这3％中近一半的交流是在布鲁塞尔－布鲁塞尔附在北部法兰德斯区域－和瓦隆之间。所以这个国家讲不同语言的两部分似乎很明确地分隔开来，只有布鲁塞尔作为它们之间的主要桥梁。



【图3】法国手机互动网络的区隔和NUTS2区域的对比



【图4】比利时手机互动网络的区隔和NUTS2区域的对比



【图5】比利时一分为二的手机互动网络和NUTS2区域的对比

最后，葡萄牙的例子同样明显。近期这个国家有很多关于如何在行政上区划葡萄牙的讨论。区域划分的问题曾经是1998年国家公投的一个话题。将我们的葡萄牙电话网络分区结果和现有的官方领土NUTS区隔作比较，不难发现第二级的分隔NUTS2（区域性政策的基本实行区域）过于粗略（其将葡萄牙领土分为5部分），而第三级的分隔NUTS3（特定区域性政策实行的小型区域）又过于细致（其将葡萄牙领土分为28部分）。我们的区隔（7部分）界于两者之间，并且和历史上的区间更为吻合，也显示出历史边界对于人类行为的影响长远，甚至超越了现代分隔的影响。国家公投时提出的区域划分和互动网络分析得到的区域划分差别更大。这大概可以解释为什么国家公投失败了。

类似的结论——基于人类互动网络得到的地理上紧密联系的社区与官方区域划分在数量和形状上的相似－在其他不同大小、不同洲的国家也都得到证实：意大利、象牙海岸、新加坡和美国。由美国电话网络推导出的社区结构与其官方州级界线间存在很多偏差。这个例子也告诉我们，在大型的、快速发展的国家里，真正的人类互动是如何越过了人为设定的、很可能已经过时了的行政区划。

虽然我们仍需要更精确的数据分析来支持城市规划的决策，然而这

个数字化分区方法显示了利用人类活动数据进行区域性研究的巨大潜力。虽然仍有诸多限制，基于电话互动网络推导的社区结构和国家区域结构之间存在普遍相关性，而这个普遍相关性的发现似乎是可靠、且能广泛应用的。所以，在基于现有方法研发出更可靠的方法之前，如果做出会影响国家区域性结构的决策或变动时，可以运用我们推导的社区结构。不同于以往用于区域性研究的调查问卷既昂贵又费时，这个新型数字化方法可以迅速，且几乎无成本地提供结果，只要能够取得一个及时更新的数据库就可以了。

注释

［1］我们在这里没有评估这些选址的统计学意义，但是有兴趣的读者可以看一下Sevtsuk的分析报告（2010）。

［2］人口资料局，http://www.prb.org

［3］欧盟统计局《NUTS分类》，http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/introduction

［4］英国统计局，http://www.neighbourhood.statistics.gov和http://www.statistics.gov.uk/2010年11月20日取得。

后记

党霄羽

《数字化城市》自我从2014年8月18日开始破土动工，至今已经过去了四个月时间。这期间，我们译者团队不仅完成了本书的翻译工作，还对本书中一些需要翻译的插图进行了图片处理，联系教授们请他们帮忙给中文版写序。多亏了优秀的队友们按时守信地交稿，对译文质量严谨负责，我们才能在四个月内完成这部10万字的译稿。

我们的团队有6个人，此前未曾谋面，大家素不相识。然而为了翻译这本书这样一个共同的目标，大家很快就成为了伙伴。队友们彼此之间的交流，无论是为了讨论翻译问题，还是请教对方的专业知识，都在非常愉快的气氛中进行。只要是和本书翻译有关的话题，即使不是自己分内的任务，大家也都非常热心地帮着出主意。我们这本书的第二个序，就是曹海元在复旦校友的名单上找到了宋朝鸣老师的联系方式，为我们争取来的。在这样一个团队中，我想其他人也都和我一样，觉得心里暖洋洋的。如今译稿已经完成，想到要和大家分别，感到十分留恋。

在此要感谢我们团队中的所有人，感谢常莹，她对数据有种执着的专业精神；感谢曹海元，他为本书中公式的呈现方式提出了建议；感谢陈然，她的校对有理有据为译者提供了方便；感谢黄艺，身在MIT的她也曾试图为本书联系一位作序者；感谢姚琪，当他发现书籍缺图的时候及时反馈，他还比其他队友多承担了一篇译文的校对工作。

此外，要代表我们整个团队感谢麻省理工大学可感城市实验室的两位主编，尽管事务繁忙，他们还是帮我们解决了本书的图片问题，还抽空给本书撰写了中文版的序言。感谢宋朝鸣老师，他为我们撰写的中文版序言中饱含着他对自己从事的专业的热爱，从中可以读出一位科学家对思考的执着与痴迷。这是一篇动人的文章，也是对后辈学子的鼓励与希冀。

我还要感谢东西文库的丁诗颖编辑对我的信任，交给我包括挑选队员、联系教授、为本书定稿等非常重要的工作。在本书翻译的过程中，她提供了许多必不可少的资料与帮助。

最后，祝愿我们国家的城市将会在各学科的研究人员的共同努力下，变得越来越宜居，也越来越人性，也希望我们这本译作在这一历史大趋势中可以起到一点小小的作用。

编辑手记：迈向智能城市

丁诗颖

工业革命之后，城市化成为全球的现象，当今中国城市化正处在一个飞速发展的阶段，如何让城市更宜居，是一个世界性的问题。《数字化城市》这本书是麻省理工大学可感城市实验室（Senseable City Lab）的一本论文集，也是东西文库的第一本公益电子书。我们希望由这本书开始，尝试纯学术领域的翻译，希望学术图书能够成为一个桥梁，让同一领域内的研究者和感兴趣的人们聚集在一起，在相互沟通中探索未来的道路。在全球化的今天，我们都生活在同一个地球上，不同国家之间的交流更为重要。希望这本书能够对中国致力于智能城市工作的创新者们有所启发，也希望通过翻译消除语言障碍，将更多有价值的研究成果展现在大众面前。

本书编者之一卡洛·拉蒂（Carlo Ratti）是麻省理工学院Senseable City Lab（可感城市实验室）的负责人，他和同事致力于让城市变得更加智能，适宜人类的居住和生活。实验室从多门学科那里获取资源，其中包括城市规划学、建筑学、设计学、工程学、计算机科学、自然科学和经济学。随着技术的革新，人与城市的对话，也在不断的变化之中。一个真正的“智能城市”是什么样子的，我们的未来的城市生活将走向何方，是我们每一个人所关心的。

作为东西文库主编，我非常荣幸这本书的中译本能够与大家见面，这将是关于智能城市的一本重要的书。感谢两位编者迪特马尔·奥夫胡贝尔（Dietmar Offenhuber）、卡洛·拉蒂（Carlo Ratti）授权译言网组织翻译本书，并愿意以互联网化的方式展现本书的成果。在本书的翻译过程中，两位编者始终给予我们各种回馈与指导，提供资料与信息，感谢他们的热情与耐心。还要感谢本书翻译团队的负责人党霄羽，以及感谢译者曹海元、常莹、陈然、黄艺、姚琪的辛勤工作。没有他们，这本书

的中译本就无法与大家见面。

让我们一起迈向智能城市，一起期待城市有一个更美好的明天。

可感城市实验室官方网站：http://senseable.mit.edu/

如果你是对本书感兴趣的研究人员，或是热心的读者，请与我们联系，我们的邮箱：dongxi@yeeyan.org